A SIMPLER, MORE GENERAL METHOD OF FINDING THE OPTIMAL FORAGING STRATEGY FOR BAYESIAN BIRDS

by

Richard F. Green

TR 2004-9

Department of Mathematics and Statistics
University of Minnesota Duluth
Duluth, Minnesota 55812

A simpler, more general method of finding the optimal foraging strategy for Bayesian birds

Richard F. Green

Department of Mathematics and Statistics, University of Minnesota-Duluth,
Duluth, Minnesota 55812 U. S. A. FAX: 218-726-8399 (e-mail: rgreen@d.umn.edu)

Abstract

Oaten's (1977) stochastic model for optimal foraging in patches has been solved for a number of particular cases. A few cases, such as Poisson prey distribution and either systematic or random search, are easy to solve. In other cases, such as binomial prey distribution and random search (Iwasa et al. 1981), the form of the optimal strategy may be found using a theorem of McNamara (1982), although more work is required to find which particular rule of the proper form is actually best. More generally (but not completely generally), optimal strategies may be found using dynamic programming. This requires that the number of prey found up to a particular time is a sufficient statistic for the number of prey remaining in a patch. This requirement cannot be dispensed with, but other simplifying assumptions that were used in the past are not necessary. In particular, it is not necessary, even for the sake of convenience, to assume that prey distribution has a form convenient for Bayesian analysis, such as a beta mixture of binomials (Green 1980) or a gamma mixture of Poissons (Green 1987a, b). Any prey distribution may be used if whatever prey are in a patch are located at random, and if search either is systematic for discrete time or for continuous time, or is random for continuous time. In earlier work (Green 1980, 1987a), some pains had to be taken to find the rate of finding prey achieved by a given candidate strategy, but this is not necessary if expected gains and expected times are calculated routinely for each potential stopping point during dynamic programming. A new, simple method of finding optimal strategies is illustrated for discrete time and systematic search. This paper is based on a talk given at the Fifth Hans Kristiansson Symposium held in Lund, Sweden in August, 2003. The subject of the symposium was Bayesian foraging.

Introduction

Animals that search for prey which are found in patches must decide when to stop searching one patch and move on to another. The problem of how long to stay in a patch, or of deciding what patch-leaving rule to use, is important to animals. Many animals live in a world in which food is distributed in patches. If some patches are much better than others, then it is important that foraging animals do not leave good patches too soon or remain in bad patches too long.

The patch residence-time problem was one of the main problems treated by optimal foraging theorists. This problem was one of four listed by Pyke, Pullian and Charnov (1977) in their review of optimal foraging theory (the others were what prey to take, what patches to search, and how to move among patches). In their definitive treatment of foraging theory, Stephens and Krebs (1986) described two basic foraging models: (1) the "patch" model, which applies to the residence-time problem, and (2) the "prey" model, which applies to the problems of choosing prey or patches.

This paper is about certain versions of the patch residence-time problem presented by Oaten (1977). In Oaten's model, foragers search for cryptic, sedentary prey, which are distributed in distinct, superficially similar patches that may vary in quality. Oaten's model is useful in the study of animal behavior and ecology. Students of animal behavior know that animals do use information obtained while foraging—that is, they behave differently depending on how many prey they have found—but it is not always clear how their experience in a patch affects their behavior. Oaten's model may be used to find how animals should use information in order to maximize their foraging rate in a particular environment.

Charnov (1976) offered a solution to the patch residence-time problem with his "marginal value theorem," perhaps the best-known idea in optimal foraging theory. Charnov's marginal value theorem says that foragers can maximize their long-term average rate of finding prey if they leave each patch when the rate of finding prey in the patch falls (perhaps due to prey depletion) to a rate equal to the long-term average rate. Actually, Charnov's theorem does not apply to predators capturing individual prey, since he assumes that foragers experience a fixed, continuous "gain function" which describes the total amount of energy gained by a forager that has spent a particular amount of time searching a patch. Charnov's theorem tells when a forager should leave a patch, but *it does not tell how an animal should decide when to leave a patch.*

While Charnov's (1976) model was deterministic, he claimed that it could easily be made stochastic, but he did not explain how to do this. Oaten (1977) developed an explicitly stochastic model to expose Charnov's claim. In Oaten's model, patches vary in the number of prey that they contain, and predators know about this variability, but they do not know how good a particular patch is before they search it. The problem that Oaten poses for foragers is how they should use the experience that they gain while searching a patch in order to decide when to leave the patch.

I have used the phrase "Bayesian birds" in the title of this paper to refer to a previous paper (Green 1980) and particularly to the dynamic programming method used there, which this paper simplifies and extends. The optimal strategy for the model foragers, which I think of as birds, is found using Bayes' theorem. The dynamic programming calculations that I have used are not the only ones possible. For example, my animals should stay in a patch at some time if there exists a strategy such that—from that time on—the ratio of expected gain (EG) to expected time to stay in a patch (ES) exceeds the highest possible long-term average rate of finding prey, C. That is, I would say: stay in a patch if EG/ES > C. McNamara (1982) defined a "potential function," which would equal EG – CES. McNamara proved a theorem showing that an optimal forager should stay in a patch if the potential function is positive, that is, if EG – CES > 0. These two formulations are mathematically equivalent, but there may be a conceptual advantage of McNamara's formulation, since the potential represents a meaningful entity, the advantage of remaining in a patch over leaving, measured in terms of the additional amount of energy to be gained during the visit. McNamara also thinks of the potential as a single entity, albeit one which is a function of the two expectations, EG and ES, while I treat EG and ES as two separate entities in my calculations. I use the EG/ES > C criterion in this paper (see equation 6).

In the following section, I describe Oaten's (1977) model, and present the formulas necessary in order to find the best patch-leaving rule and the rate that it achieves. Then the method is used to find the best rule and the rate that it achieves for an experiment performed by Lima (1984) on Downy woodpeckers. Finally I discuss the role that Oaten's model has played and, I hope, will continue to play in the study of foraging.

An annotated Turbo Basic program that does the discrete-time calculations is included as an Appendix.

Oaten's model

Allan Oaten (1977) proposed an explicitly stochastic model to expose Charnov's (1976) claim that his deterministic "marginal value theorem" could be extended easily to a stochastic model. The general question is: if a forager searches for prey which are found in distinct patches, what rule the forager should use to decide when to leave one patch and go on to the next? In particular, if patches vary in quality, how does a forager decide that the present patch is poor and should be abandoned?

Assumptions of Oaten's model

Oaten (1977) envisioned a forager that searches superficially similar patches that may vary in quality. He assumed:

1. Prey are found in patches and the predator knows the distribution of the number of prey per patch.

2. The predator knows the joint distribution of the capture times, given the number of prey in a patch.

3. Prey are not replaced as they are captured.

4. The predator knows the time τ that it takes to travel from one patch to the next.

5. The predator decides when to leave a patch based on its knowledge of 1, 2, and 4 and its experience in the patch.

6. Given a strategy, we can calculate

$$R = EG/(ES + \tau),$$

where

EG = the average number of prey caught per patch,
ES = the average length of time per patch, and
R = the long-term average rate of gain, using the given strategy.

7. The predator uses the strategy that maximizes R.

Several comments might be made about Oaten's model.

a. Unlike the deterministic model of Charnov (1976), Oaten's model does not pretend to be general. It applies to a number of particular cases, but these cases must be treated separately. Charnov's model is not truly *general*. That is, it does not *apply* to *many* cases. Rather, it is *generic*, in the sense that it *refers* to *no* particular case. If, in fact, different animals face different foraging problems, calling for different solutions, then it is useful to have a theory that has a place for different cases, leading to different results.
b. Oaten's model makes explicit assumptions that have biological meaning. For example, patches do have varying numbers of prey, and the model has a place for the distribution of prey numbers. Given a particular number of prey in a patch, the times at which they will be encountered depend on the pattern of search—for example, it might be systematic, or it might be random.
c. Despite the fact that Oaten's model applies to a variety of particular cases, it suffers from some limitations. It assumes that patches are superficially similar (that is, they are all of the same "type"). This point was offered as a criticism of Oaten's model by Stephens and Charnov (1982), who distinguished between patch "types" (which can be distinguished upon inspection, without search) and patch "sub-types," (which can only be distinguished by being searched). Prey are also assumed to be sedentary and cryptic in Oaten's model.

d. Oaten's model is very complicated in principle, and only relatively simple examples have been worked out. The examples that I have worked out require particular assumptions about prey distribution and pattern of search (assumptions 1 and 2 above). The point of this paper is to show how restrictions on assumption 1 may be removed.

What it looks like in a patch

A patch may be visualized as a line segment. This may be thought of as a time line, which may be assumed, without loss of generality, to run from time, $t = 0$ to $t = 1$ for continuous time, or from $t = 0$ to $t = n$ for discrete time. The times at which prey are captured are points on the time line. The experience of a forager in a patch may be represented by a function of time, $X(t)$, which is the number of prey which have been captured by time t. For Oaten's model, $X(t)$ is a stochastic process. The stochastic process, $X(t)$, corresponds to Charnov's (1976) deterministic function, $g(t)$, which he calls the "gain function."

The problem that a forager faces in Oaten's model is to decide what to do if it finds itself at point $(t,X(t))$, that is, if it has been in a patch for time = t and has found X prey by that time. Of course, what happens after time t depends on chance. This presents a complication, but it is not too difficult to handle mathematically. A more serious difficulty is that the situation at time t may depend on the sequence of capture times before time t and not just on the number of prey that have been captured by time t. In order to eliminate this difficulty, I add restrictive assumptions to Oaten's so that the number of prey found by time t is a sufficient statistic for the number of prey in the patch. [This point was mentioned by Iwasa *et al.* (1981)]. That is, I make Oaten's assumption 2 more restrictive.

In the discrete time case that I consider in this paper, I get the sufficiency that I want by assuming that search is systematic and that whatever prey are in the patch are located at random. More generally, the number of prey found by time t will be a sufficient statistic for the number of prey in the patch if prey are located at random and if the proportion of the patch searched by time t is a deterministic function of t. This latter condition is satisfied for systematic search, whether time is discrete or continuous, and if search is random and time is continuous. However, if search is random and time is discrete, the proportion of a patch searched by time t is unpredictable, and the number of prey found by time t is not sufficient for the number of prey in a patch. I have not treated this case in any of my work.

It is natural to look at a forager's experience in a patch as the number of prey captured as a function of time. Another method is to look at the instantaneous rate of finding prey (for continuous time, or the probability of finding a prey in the next bit searched for discrete time) as a function of time for a forager that has certain experience in a patch. Iwasa *et al.* (1981) illustrated this function in their Fig. 4., and they suggested that a forager might best leave a patch when the value of this function fell to some critical threshold. A rule such as that suggested by Iwasa *et al.* is sometimes best, but it is not

always best. More generally, it is necessary to consider a forager that has found x prey by time t and to think about what would happen to such a forager *for the rest of its time in the patch* and not just what would happen in the next bit of the patch. McNamara (1982) made this point clear. One way of understanding McNamara's idea is that if patches vary, then a forager gains not only prey, but also information about patch quality as it searches a patch. A forager may stay in a patch that seems poor, especially at the beginning, in order to gain information about its quality

Method

I have found the optimal foraging strategy for a number of examples of Oaten's model using dynamic programming (Green 1980, 1987a, 1988b). Each of the examples was chosen because it yielded a simple form for the posterior distribution of the parameter of interest [binomial p (Green 1980, 1988b), Poisson λ (Green 1987a)]. The point of this paper is to show how to find the optimal strategy for any prey distribution. In my published work, I found a candidate optimal patch-leaving rule and the rate that it would achieve in two separate operations. Here I illustrate a method used in an unpublished technical report (Green 1988b) that finds candidate rules and the rates that they achieve in one operation.

Discrete time

Basic ideas

A forager that searches for prey distributed in patches may leave a patch at any time, and *from the time that it leaves a patch it may achieve a long-term average rate equal to R*, the highest possible long-term average rate.* Therefore, the only reason for staying in a patch is that it is possible to achieve at least as high a rate *for the rest of the time spent in the patch*. A forager should leave a patch when it reaches point (t,x) unless there is some rule such that, from this point on, the ratio of expected gain to expected time in the patch is at least R*. That is, a forager should stay in a patch if there exists a rule such that

$$E[G(t,x)]/E[S(t,x)] > R^*, \qquad (1)$$

where E[G(t,x)] and E[S(t,x)] are the expected gain and the expected time in the patch from point (t,x) on, respectively, for a forager using the rule. Or course, we do not know R* or the desired strategy. It is our job to find them.

The idea is to guess a value for R*, call it C, and find a patch-leaving rule that "tries to achieve" rate C, that is, a rule that would achieve rate C, if, in fact, C were the highest possible rate. The candidate best rule for a guess C, and the rate that it actually achieves, R(C), are found using dynamic programming, beginning at the end of the patch and working backward to the beginning.

Using dynamic programming

If each patch consists of n discrete bits, each of which might contain a prey, we begin at time $t = n - 1$, when exactly $n - 1$ bits have been searched. We calculate the expected gain and the expected time in the remainder of the patch *assuming that the forager decides to go on and search the final bit.* For each possible number of prey found, x, we denote the expected gain and expected time as $g(t = n - 1,x)$ and $s(t = n - 1,x)$, respectively. We have $g(n - 1,x) = r(n - 1,x)$, the probability that the last bit will contain a prey, given that x prey have been found in the first $n - 1$ bits, and $s(n - 1,x) = 1$, since we assume that the forager will go on and search one more bit.

Once we have found $g(n - 1, x)$ and $s(n - 1, x)$ we can determine whether to stay by checking whether

$$g(n - 1,x)/s(n - 1,x) > C. \tag{2}$$

If this inequality is satisfied, then the forager should stay in a patch and search the last bit when x prey have been found by time $t = n - 1$, otherwise the forager should leave the patch at time $t = n - 1$. In fact, it does not matter what the forager decides if the equality holds, but here, for definiteness, I decide that the forager should stay if and only if the inequality in (2) is satisfied. If (2) is satisfied one can set $G(n - 1,x) = g(n - 1,x)$ and $S(n - 1,x) = s(n - 1,x)$, where $G(n - 1,x)$ and $S(n-1,x)$ are, respectively, the expected gain and the expected time for the rest of a patch for a forager finding itself at point $(n - 1,x)$ and using a strategy that "tries to achieve" rate C.

After we know what to do at time $t = n - 1$, for each possible number of prey, x, found by that time, we can go backwards, finding what to do at time $t = n - 2$, then $t = n - 3$, and so on, until $t = 0$. When we reach $t = 0$ we will have

$$R(C) = G(0,0)/[S(0,0) + \tau], \tag{3}$$

where R(C) is the long-term average rate of finding prey achieved by a forager that "tries to achieve" rate C. The simplicity of equation (3) and the argument that leads to it is the first of the two main mathematical points of this paper.

For general time t and number of prey found x, we *tentatively assume that the forager will go on and search bit t + 1, and after that time it will use the rule that has been built up by working backward to that time.* The expected gain and expected time in the remainder of the patch, *under that assumption* are given by

$$g(t,x) = r(t,x)\,[1 + G(t + 1,x + 1)] + [1 - r(t,x)]\,[G(t + 1,x)], \text{ and} \tag{4}$$

$$s(t,x) = 1 + r(t,x)\,[S(t + 1,x + 1)] + [1 - r(t,x)]\,[S(t + 1, x)], \tag{5}$$

respectively, where $r(t,x)$ is the probability that there will be a prey in the $t + 1^{st}$ bit searched, given that x prey have been found in the first t bits searched.

In order to decide whether to go on and search the next bit if x prey have been found by time t, we check

$$g(t,x)/s(t,x) > C\,. \tag{6}$$

Now we can define a "decision function," d(t,x), which takes a value of 1 if (6) is satisfied, and a value of 0 if (6) is not satisfied. Then we have

$$G(t,x) = d(t,x)\, g(t,x), \text{ and} \tag{7}$$

$$S(t,x) = d(t,x)\, s(t,x), \tag{8}$$

for the expected gain and expected time, respectively, for the rest of the patch for a forager that reaches point (t,x) and "tries to achieve" rate C.

Finding the probability r(t,x)

The second main mathematical point of this paper is to show how to find r(t,x) for a general distribution, f(y), of the number of prey, y, in a patch. In earlier work (Green 1980, 1988b) the prey distribution was chosen so as to make the calculation of r(t,x) easy. For example, if the number of prey in a patch has a binomial distribution with parameters n and p, but p itself has a beta distribution [the "conjugate prior" of the binomial distribution] with parameters α and β, then a standard Bayesian analysis (see DeGroot 1970) shows that if x prey have been found by time t, the posterior distribution of p has a beta distribution with parameters $\alpha + x$ and $\beta + t - x$. This produces a very simple form for r(t,x):

$$r(t,x) = (\alpha + x)/(\alpha + \beta + t). \tag{9}$$

Several examples of this case were treated by Green (1980). Another particular case, in which the prior distribution is uniform [a beta distribution with parameters $\alpha = 1$ and $\beta = 1$] has been treated by Xu (2000) and by Riley (2003). For this particular case, r(t,x) has the form:

$$r(t,x) = (x + 1)/(t + 2).$$

There is an attractive simplicity in choosing prey distributions for which the calculations are easy. However, experimenters sometimes choose prey distributions which are easy for them to use, even though the model would be more difficult to treat analytically [for examples, see Lima (1984), Mellgren, Misani and Brown (1984), Mellgren and Brown (1987)]. In fact, experimenters tend not to analyze their models mathematically, so the question of whether their models are mathematically tractable may seem moot. However, a careful analysis of an experimental model may show things that the experimenter did not consider.

Here I will consider a general distribution, f(y), for the number of prey, y, that a patch contains. If, by time t, x prey have been found in a patch of size n *that begins with y prey*, then the probability of finding a prey in the next bit searched will be

$$(y - x)/(n - t). \tag{10}$$

The number of prey in a patch, Y, is a random variable, with distribution f(y). If x prey have been found by time t, we denote the conditional distribution of the number of prey in the patch y, given x, by f(y | x), and then we can find r(t,x) by averaging (y – x)/(n – t) over the conditional distribution of y given x. That is,

$$r(t,x) = \Sigma\, f(y \mid x)\, (y - x)/(n - t). \tag{11}$$

The conditional distribution of y given x is found using Bayes' theorem:

$$f(y \mid x) = f(y)\, p(x \mid y)/[\Sigma\, f(y)\, p(x \mid y)], \tag{12}$$

where p(x | y) is the conditional probability that X = x given that Y = y and has a hypergeometric distribution. That is,

$$p(x \mid y) = C(t,x)\, C(n - t, y - x)/C(n,y), \tag{13}$$

where C(n,y) = n!/[y!(n – y)!] is the number of combinations of n things taken y at a time.

Continuous time

In continous-time models it is assumed that prey can be found at any time, not just at discrete times. For systematic search, the forager may be envisioned as moving through a patch smoothly, at a constant rate. I assume that whatever prey there are in a patch are located at random, so a systematic forager is equally likely to find prey at the beginning or the end of a patch visit. For systematic search, each patch is assumed to have a distinct end, and the forager leaves a patch after reaching the end. Without loss of generality, it may be assumed that each patch is of size one, which is equivalent to assuming that it takes unit time to search each patch. Under these assumptions, a forager that has searched a particular patch for time t will have searched exactly proportion t by that time. If the patch began with y prey, a forager that has searched for time t will have found a random number of prey, X, which will have a binomial distribution with parameters n = y and p = t. For random search it is assumed that unsearched portions of a patch will be visited at an exponentially decreasing rate, and without loss of generality, it may be assumed that the rate of decrease equals one. This implies that a random forager that has spent time t in one patch will have searched exactly proportion 1 – exp(-t) of the patch by that time. If a patch begins with y prey, the number of prey, X, found by a random forager that has spent time t in the patch will have a binomial distribution with parameters n = y and p = 1 – exp(-t).

a) Systematic search

The optimal strategy for systematic search and continuous time may be found approximately by using the method outlined above for discrete time by letting a patch consist of many bits in which prey may be found. This method may be modified by replacing the hypergeometric distribution in (13) with a binomial distribution and by replacing expressions (4) and (5) for g(t,x) and s(t,x) respectively by sums over the number of prey that might be found in the next bit [which need not be confined to 0 or 1, as it is for discrete time]. I have not explored this modification of the discrete-time model for systematic search, but slightly more extensive modifications are needed for random search, and these are discussed in the next section.

b) Random search

Random search for continuous time differs from systematic search for discrete time in several ways.

(1) For random search there is no end of a patch. However, continuous time random search may be viewed as systematic search at an exponentially decreasing rate. Without loss of generality, I may assume that a patch is searched at unit rate. This means that if search were systematic, the entire patch could be searched in unit time. For random search, the proportion of a patch that is searched by time t is exactly $1 - \exp(-t)$. For example, by time $t = 3$, proportion $1 - \exp(-3)$ of a patch will have been searched. Perhaps more to the point, only proportion $\exp(-3) \approx 0.05$ will remain unsearched and most of the prey originally in the patch will have been found.

(2) With random search, the future success of a forager that has spent some time, t, in a patch depends only on the number of prey, $y - x$, remaining in the patch at that time. Given the value of $y - x$, future success does not depend on the time in the patch, t, as it did for systematic search, as shown in expression (11). The consequence of this will be seen in expression (17) below.

(3) It is possible that two or more prey are found in an interval of length Δ. This implies that (4) and (5) might be replaced by sums:

$$g(t,x) = \Sigma\, r(i, t, x)\, [i + G(t + \Delta, x + i)], \text{ and} \tag{14}$$

$$s(t, x) = \Delta + \Sigma\, r(i, t, x)\, [S(t + \Delta, x + i)], \tag{15}$$

where $r(i, t, x)$ is the probability that exactly i prey will be found in the next time interval of length Δ, given that x prey have been found by time t.

If, at any time, the number of prey remaining in a patch equals $y - x$, then the distribution of the number of prey, I, caught in the next interval of length Δ will have a binomial distribution with parameters $n = y - x$, and $p = \Delta_1 = 1 - \exp(-\Delta)$. That is, the distribution of I when $y - x$ prey remain in the patch is given by

$$b(i, y - x) = C(y - x, i)\, \Delta_1^{\,i}\, (1 - \Delta_1)^{y-x-i}. \qquad (16)$$

Then we will have r(i, t, x) given by

$$r(i, t, x) = \Sigma\, f(y \mid x)\, b(i, y - x). \qquad (17)$$

(4) The distribution of X(t), the number of prey that are to be found by time t in a patch, given that the patch contains y prey, is binomial with parameters n = y and p = 1 – exp(-t). That is,

$$p(x \mid y) = C(y, x)\, [1 - \exp(-t)]^x\, [\exp(-t)]^{y-x} \qquad (18)$$

The optimal strategy for continuous time and random search is found by dynamic programming using the same method as used for discrete time, except that for random time one starts after some long time (T = 3 or 4, say) in the patch and works backward in steps of length Δ (= 1/N = 1/100, say). The steps in the calculation for continuous time are the same as for the discrete time case, and some of the same formulas are used [(1), (3), (6), (7), (8) and (12) are maintained, while (2) is replaced by

$$g(T - \Delta, x)/s(T - \Delta, x) > C \qquad (19)$$

and (4), (5), (10), (11) and (13) are replaced by (14), (15), (16), (17) and (18), respectively.]

Results

The method is illustrated for one discrete-time example, a model used by Lima (1984) in an experiment on Downy Woodpeckers. Patches were artificial "trees"—wooden dowels, each with 24 holes in six rings of four holes drilled around the dowel. In one experiment, sixty percent of the patches were "empty"—with all the holes empty—and forty percent of the patches were "full" (actually, half-full)—with exactly twelve holes having "prey" (bits of sunflower seed).

The birds searched the patches systematically, and for systematic search, having an exact number of prey causes some difficulty. It would be much simpler computationally to have "full" patches with random numbers of prey, where the distribution was binomial with n = 24 and p = 0.5. Then "full" patches would not all have exactly twelve prey, but the *average* for "full" patches would be $\mu = np = 24(0.5) = 12$ prey. In that case, the optimal strategy would be to remain in each patch until some fixed number, k, of empty holes had been encountered, and then leave. But, if a "prey" is found before k empty patches have been searched, the bird should remain in the patch and search it completely. (I refer to this simple rule as "Lima's rule.") This simplified version of Lima's model has been discussed by Stephens and Krebs (1986) and by Green (1990).

[Put Figure 1 here.]

The best rule for Lima's "half-full" experiment is shown in Figure 1. (Here I assume that travel time is $\tau = 4$, which seems to be close to what it was in the experiment.) The best rule is uglier than "Lima's rule" because the bird should leave a patch, not only when k (here k = 3 is best) empty holes have been searched at the beginning of a patch visit, but also if many prey are found quickly—for example, if 9 prey are found in the first 9 or 10 holes searched, or if 10 prey are found by the time 14 holes have been searched—then the bird should leave the patch. The points marked with dots are points that the forager might reach, and for which it should continue searching the patch. The points marked with X are stopping points, or points that should not be reached because the forager should have left the patch before reaching them.

Theory shows the form of the best patch-leaving rule and the rate of finding prey that a forager using that rule would achieve. For a given model, rates may be calculated for other rules, perhaps including the rule that animals actually use. For the example of Lima's "half full" model considered here, the naïve rate that would be obtained by a forager that ignored information and searched each patch completely would be r = 0.1714, which is about 56% of the optimal rate, r = 0.302784, that would be obtained by an animal using the rule illustrated in Figure 1. It may be implausible to assume that an animal could count up to twelve prey, so we might consider the simple rule, which I refer to as "Lima's rule": leave a patch if the first k holes searched are empty, otherwise stay until the patch has been searched completely. Here the best k = 3, and the rate achieved, r = 0.295318, is almost 98% of that achieved by the best rule. In fact, the animals in Lima's experiment did not search a fixed number of empty holes before leaving a patch, but sometimes left after k = 1, sometimes after k = 2, and so on. If an animal always left a patch after finding a particular number of prey, k = 2, say, one could calculate the rate that such a forager might achieve. If we think of the actual behavior of the animal as consisting of a mixture of pure rules (with a certain probability use k = 1, with a certain other probability use k = 2, and so on), then we can calculate the rate for each possible value of k and then find the average these rates over the frequency distribution observed in the experiment. [This idea is illustrated in Green (1990, Fig. 4).] If this averaged value is used, the observed rate is r = 0.2859, which is more than 94% of the optimal rate.

What these calculations show is that in Lima's experiment the animals can find prey at a much higher rate if they use information about patch quality than if they do not use information. Further, if we calculate the rate actually achieved by the animals, we find that it also is much greater than the naïve rate and almost as high as the maximum rate. Lima did not do these calculations. He did comment on the substantial variability in the observed rule. He found that the coefficient of variation (σ/μ) for k, the number of empty holes searched before leaving a patch, was 0.4284. While this might seem like quite a large relative variation in behavior, calculating the rates actually achieved shows that the animal does nearly as well as possible despite the variability in behavior. The reason that this happens is that the rate of finding prey is not very sensitive to the value of k used. The important point is that the quality of foraging behavior should be measured by the

rate of finding prey that it achieves, not by how similar the actual behavior is to the behavior that would achieve the highest rate.

The optimal strategy for Lima's model was calculated using dynamic programming, using the simplified, general method described in this paper. Other examples have been calculated using a more complicated method, which I have described earlier (Green 1980, 1987a). Several examples have been illustrated in the same way as here, with a figure showing the best patch-leaving rule, and a comparison of the rates of the optimal rule, the naïve rule, and some other rules. A discrete-time example is given in Green (1980) and several continuous-time examples are given is Green (1987a). More mathematical details of the old method are to be found in two unpublished technical reports (Green 1987b for systematic search and Green 1988a for random search). These technical reports are available to interested readers upon request from the author.

Discussion

In studying animal behavior we may ask: What problem does an animal face? How does it solve the problem? And, what are the consequences of the solution? The problem of interest in this paper is how an animal decides when to leave a patch if it forages in a environment in which prey are distributed in patches that differ in quality. The approach is to use a mathematical model of the foraging problem and to determine the best solution to that problem. Oaten's (1977) stochastic model is valuable because it requires an explicit statement of the problem that a forager faces: how prey are distributed in the environment and how foragers may search for them. I favor a quantitative treatment of particular cases. This makes it possible to compare the performance of different foraging strategies for each case, and to study the population consequences of particular foraging strategies in particular cases.

Early work on how a forager decides when to leave a patch was flawed because it did not have the advantage of Oaten's theory. For example, in their laboratory study of Black-capped chickadees, Krebs, Ryan and Charnov (1974) compared three rules that a bird might use to decide when to leave a patch, and they performed an experiment to determine which rule the birds actually use. Unfortunately, Krebs, Ryan and Charnov did not know what prey distribution their birds faced in nature, they did not know what the optimal solution was for the problem that they set for their birds, and they did not know how to analyze the data obtained from their experiment (see Green 1987a for an elaboration of these points).

Hassell and May (1974) studied the population consequences of foraging in a predator-prey model, but they did not calculate the foraging rate properly. Murdoch and Oaten (1975) also studied the consequences of foraging for a predator-prey model, using the same foraging rule as Hassell and May (1974) (the fixed giving-up time rule), and they calculated the rate correctly. However, Murdoch and Oaten (1975) overstated the generality of conditions permitting stability because the giving-up time rule is not optimal for either of the prey distributions that they considered, and the giving-up time that they used in their calculations was better for average prey densities larger than the equilibrium

values whose stability was being examined. The model foragers seemed to feed more efficiently at higher prey densities (and thus produce a system-stabilizing sigmoid functional response) because their foraging became less inefficient at higher prey densities. If predators forage optimally then the conditions for stability are more limited than Murdoch and Oaten (1975) believed (Green 1990).

After Oaten's (1977) paper was published there was a spurt of work on particular examples of his model (Breck 1978; Green 1980; Iwasa et al. 1981; McNamara 1982). Breck (1978) and Iwasa et al. (1981) considered several possible prey distributions and compared the performance of various rules for each. Green (1980) and McNamara (1982) found optimal strategies for particular cases. The fixed giving-up time rule was best among the rules considered by Breck (1978) and by Iwasa et al. (1981) for a negative-binomial prey distribution. McNair (1982) showed that the fixed giving-up time rule was not generally best, as Krebs, Ryan and Charnov (1974) had claimed. Later Green (1987a) showed that the giving-up time rule is not optimal for a negative binomial prey distribution, but that it is quite good, at least for the parameters that he used. Green (1980, 1987a) used dynamic programming to find the optimal patch-leaving rules for a number of biologically important prey distributions, but this method was not used often for Oaten's (1977) model until the recent work of Olsson and Holmgren (1998, 2000).

Not much work, either theoretical of experimental, was done on Oaten's (1977) model for some time after the initial burst of activity. Stephens and Krebs (1986) mention Oaten's model, and they criticize Oaten and his followers for ignoring the issue of distinguishable patch types, which Charnov (1976) had raised. The issues that Oaten had raised, the questions of how prey are distributed in patches and how foragers use information to decide when to leave each patch, did not attract as much interest among experimentalists as the simpler basic foraging questions of whether foragers are more selective of prey when more good prey are available, or whether foragers stay longer in a patch when the travel time to the next patch is greater. Lima (1984, 1985) did experiments on how foragers decide when to leave patches in a variable environment, including the experiment on downy woodpeckers treated theoretically in this paper, and an experiment on starlings.

Haccou et al. (1991) used reliability theory to study the rule that a parasitic wasp uses to decide when to leave a patch of hosts. They observed the time from the last host encounter until a wasp left a patch (the giving-up time). The giving-up time is variable, and its average depends on the number of hosts discovered in a patch. Wildhaber et al. (1994) studied the giving-up time of bluegill sunfish offered discrete bits of food. They showed that the average giving-up time increased with the number of food bits obtained in a single foraging bout. Van Gils et al. (2003) considered the effect of digestive constraints on the foraging efficiency of red knots in an experiment that offered varying numbers of difficult-to-digest prey in artificial patches.

Eventually, work on Oaten's (1977) model resumed, largely done by people who participated in the 2003 Bayesian foraging symposium at Lund. Tom Valone and colleagues (Valone 1989, 1991; Valone and Brown 1989; Valone and Giraldeau 1993)

studied information use in foragers, or in groups of foragers. More recently, Sernland et al. (2003) studied the effect of information sharing on group foraging. Group foraging adds a number of issues that are not considered by students of solitary foragers, including the question of whether individuals can do better in groups than in solitude, and, if they forage in groups, whether they can use information about patch quality that is obtained by others, and if so, how. A recent book by Giraldeau and Caraco (2000) deals with the many issued involved in group foraging.

Rodriguez-Girones and Vasquez (1997) looked in some detail at the results of Bayesian foraging for several different prey distributions. They were interested in what they referred to as density dependence, that is, the relationship between the proportion of prey taken from a patch and the number of prey originally in the patch. They found that if prey are aggregated, in particular, if they come from a negative binomial distribution, then an optimal forager would take a larger proportion of prey from patches beginning with more prey. That is, harvesting should be positively density dependent. If prey have a Poisson distribution, then an optimal forager would take the same proportion of prey from each patch and harvesting would be density independent. If prey distribution has less variability that the Poisson, then an optimal forager would take a higher proportion of prey from patches with fewer prey, and harvesting would be negatively density dependent. Perhaps it should be mentioned that when population ecologists talk about density dependence they are referring to a larger scale than Rodriguez-Girones and Vasquez (1997) have in mind. Population ecologists refer to prey density within an environment, not within patches in an environment. If we are interested in the impact of foraging on the stability of a predator-prey system (see for example, Murdoch and Oaten 1975; Green 1990), the conditions for stability involve differences in foraging rate between different environmental conditions, not in the proportions of prey taken from different patches within a particular environment.

Olsson and Holmgren (2000) criticized Rodriquez-Girones and Vasquez (1997) by showing that they had used a sub-optimal patch-leaving rule. Rodriguez-Girones and Vasquez assumed that optimal foragers should leave a patch when the expected instantaneous rate of finding prey falls to a fixed, critical value. This is a mistake that has been made several times in the history of optimal foraging theory (for examples, see Pyke 1978 and Iwasa et al. 1981). Olsson and Holmgren used dynamic programming to find the optimal strategy and showed that for a negative binomial prey distribution the instantaneous rate of finding prey at particular patch-leaving points increases with the number of prey found. That is, the longer a forager has spent in a patch the better opinion it must have about patch quality in order to stay. This fact means that optimal foragers should not show as strong density dependence as would foragers using the instantaneous rate rule. Olsson and Holmgren did their calculations making the usual assumption that optimal foragers maximize their long-term average rate of finding prey, but they also considered predation minimization, in which the optimization criterion is to minimize the forager's risk of predation while it forages. Foragers that are reluctant to leave patches because of increased predation risk while traveling will tend to stay longer in each patch, and the density dependence is not as pronounced as it is for rate maximizers.

Dynamic programming is a method that is needed to find the optimal patch-leaving rule for particular examples of Oaten's (1977) general stochastic foraging model. Dynamic programming has been used relatively few times for Oaten's model (Green 1980, 1987a; Olsson and Holmgren 1998, 2000), but it has become a popular method under the heading of SDP, or "stochastic dynamic programming," for modeling behavioral decisions. This method emerged in the mid-1980s (Mangel and Clark 1986; McNamara and Houston 1986) as a powerful method for finding optimal behavioral policies for animals living under changing conditions. One of the first problems involved foraging: how should a small bird allocate its time during a winter day in order to maximize its chance of obtaining enough food to survive the night? Such an animal was assumed to face a trade-off between the need for food and the risk of predation. The optimal behavior depended on the time of day and the nutritional state of the bird. The paper by Mangel and Clark (1986) purported to develop a "unified foraging theory," but what it actually did was combine a single, simple foraging problem with other problems not involving foraging, and showed how to solve them all simultaneously. Mangel and Clark did not offer a unified treatment of foraging theory.

Since stochastic dynamic programming emerged as a method of modeling behavioral problems, there have been two articles published as target papers in *Behavioral and Brain Sciences* (Houston and McNamara 1988; Clark 1991), three books (Mangel and Clark 1988; Houston and McNamara 1999; Clark and Mangel 2000), and many journal articles. There is a philosophical difference in the way that dynamic programming is used to find the optimal patch-leaving rule for Oaten's model and the way that it is used in SDP. Work on foraging theory in general, and Oaten's model in particular, is problem-driven. Workers use whatever method is appropriate to solve each part of a related set of problems. For example, in order to compare the performance of various rules, dynamic programming is used only to find and evaluate the optimal rule, while other methods are used to evaluate other rules. Work using SDP tends to be method-driven. Workers who know how to use the method seek problems, or parts of problems, for which the method provides the answer. There are many such problems, and SDP has been a useful method. However, foraging theory may have suffered from the success of SDP because people who had the technique and the ability to solve problems in optimal foraging used their ability to solve a lot of other problems instead.

Appendix: annotated Basic program

Below is a Turbo Basic program that was used to find the optimal patch-leaving rule for an example of Oaten's model: Lima's (1984) "half-full" experiment on Downy Woodpeckers.

Lines 100-110 specify the prey distribution, f(y).

Lines 200-230 calculate log n!

Lines 300-340 calculate log C(i, j).

Line 350 specifies the "guessed rate," C—the best rate (found earlier) is used.

Lines 435-540 determine the sums for the denominator needed in (12), which is needed for (11), and the numerator for (12).

Line 620 finishes the calculation of r(t,x).

Lines 630-640 find g(t,x) and s(t,x) for (14) and (15).

Line 660 tests the rate as indicated in (6).

Lines 670-680 find G(t,x) and S(t,x) for (7) and (8).

Lines 910-980 find and print the stopping rule.

[Put the Basic program here.]

```
10 Print "Lima for Lund"

20 n = 24
30 t0 = 4
40 Dim c(n, n), r(n, n), d(n, n), g1(n, n), s1(n, n), g(n, n), s(n, n), f(n), l(n)

100 f(0) = 0.6
110 f(12) = 0.4

200 l(0) = 0
210 For i = 1 To n
220 l(i) = l(i - 1) + Log(i)
230 Next i

300 For i = 0 To n
310 For j = 0 To i
320 c(i, j) = l(i) - l(j) - l(i - j)
330 Next j
340 Next i

350 For c = 0.302784 To 0.31
```

```
400 For i = 0 To n
410 g(i, n) = 0
420 s(i, n) = 0
430 Next i

435 d1 = 0
436 n1 = 0

440 For t = 23 To 0 Step -1

450 For x = 0 To t
460 n1 = 0
470 d1 = f(x) * Exp(c(x, x) + c(n - x, t - x) - c(n, t))

500 For y = x + 1 To x + n - t
510 w = c(y, x) + c(n - y, t - x) - c(n, t)
520 d1 = d1 + f(y) * Exp(w)
530 n1 = n1 + f(y) * Exp(w) * (y - x) / (n - t)
540 Next y


600 r(t, x) = 0
610 If d1 = 0 Then GoTo 630
620 r(t, x) = n1 / d1
630 g1(t, x) = r(t, x) * (1 + g(t + 1, x + 1)) + (1 - r(t, x)) * g(t + 1, x)
640 s1(t, x) = 1 + r(t, x) * s(t + 1, x + 1) + (1 - r(t, x)) * s(t + 1, x)
650 d(t, x) = 0
655 If s1(t, x) = 0 Then GoTo 670
660 If g1(t, x) / s1(t, x) > c Then Let d(t, x) = 1
670 g(t, x) = d(t, x) * g1(t, x)
680 s(t, x) = d(t, x) * s1(t, x)

700 Next x

800 Next t

900 Print using; "##.######  "; c; g(0, 0); s(0, 0); g(0, 0) / (s(0, 0) + t0)

910 For t = 1 To n
920 For x = 1 To t - 1
930 If d(t, x) > d(t, x + 1) Then Print t, x + 1
940 Next x
950 Next t

Print

960 For t = 0 To n - 1
970 If d(t, 0) > d(t + 1, 0) Then Print t + 1, 0
980 Next t


1000 Next c
```

References

Breck, J. E. 1978. Suboptimal foraging strategies for a patchy environment. – Ph. D. thesis, Michigan State University.

Charnov, E. L. 1976. Optimal foraging: the marginal value theorem. – Theoretical Population Biology 9, 129-136.

Clark, C. W. 1991. Modeling behavioral adaptations. – Behavioral and Brain Sciences 14, 85-117.

Clark, C. W. and Mangel, M. 2000. Dynamic State Variable Models in Ecology. – Oxford University Press, Oxford.

DeGroot, M. H. 1970. Optimal Statistical Decisions. – McGraw-Hill, New York.

Giraldeau, L.-A. and Caraco, T. 2000. Social Foraging Theory. – Princeton University Press, Princeton, New Jersey.

Green, R. F. 1980. Bayesian birds: a simple example of Oaten's stochastic model of optimal foraging. – Theoretical Population Biology 18, 244-256.

Green, R. F. 1984. Stopping rules for optimal foragers. – American Naturalist 123, 30-43.

Green, R. F. 1987a. Stochastic models of optimal foraging. – In: Foraging Behavior. (Kamil, A. C., Krebs, J. R. and Pulliam, H. R., eds.) Plenum Press, New York, pp. 273-302.

Green, R. F. 1987b. A stochastic model of optimal foraging: systematic search for negative-binomially distributed prey. – University of Minnesota-Duluth, Department of Mathematics and Statistics, Technical Report 87-2.

Green, R. F. 1988a. Optimal foraging for patchily distributed prey: random search. – University of Minnesota-Duluth, Department of Mathematics and Statistics, Technical Report 88-2.

Green, R. F. 1988b. Optimal foraging in Bernoulli patches which are of two types. – University of Minnesota-Duluth, Department of Mathematics and Statistics, Technical Report 88-4.

Green, R. F. 1990. Putting ecology back into optimal foraging theory. - Comments on Theoretical Biology 1, 387-410.

Haccou, P., de Vlas, S. J., van Alphen, J. J. M. and Visser, M. E. 1991. Information processing by foragers: effects of intra-patch experience on the leaving tendency of *Leptopilina heterotoma*. – Journal of Animal Ecology 60, 93-106.

Houston, A. I. and McNamara, J. M. 1988. A framework for the functional analysis of behavior. – Behavioral and Brain Sciences 11, 117-163.

Houston, A. I. and McNamara, J. M. 1999. Models of Adaptive Behaviour: An Approach Based on State. – Cambridge University Press, New York.

Hassell, M. P. and May, R. M. 1974. Aggregation of predators and insect parasites and its effect of stability. – Journal of Animal Ecology 43, 567-594.

Iwasa, Y., Hagashi, M. and Yamamura, N. 1981. Prey distribution as a factor determining the choice of optimal foraging strategy. – American Naturalist 117, 710-723.

Krebs, J. R., Ryan, J. C. and Charnov, E. L. 1974. Hunting by expectation or optimal foraging? – Animal Behaviour 22, 953-964.

Lima, S. L. 1984. Downy woodpecker foraging behavior: efficient sampling in simple stochastic environments. – Ecology 65, 166-174.

Mangel, M. and Clark, C. W. 1986. Toward a unified foraging theory. – Ecology 67, 1127-1138.

Mangel, M. and Clark, C. W. 1988. Dynamic Modeling in Behavior Ecology. – Princeton University Press, Princeton, New Jersey.

McNair, J. N. 1982. Optimal giving-up times and the marginal value theorem. – American Naturalist 119, 511-529.

McNamara, J. 1982. Optimal patch use in a stochastic environment. – Theoretical Population Biology 21, 269-288.

McNamara, J. M. and Houston, A. I. 1986. The common currency for foraging decisions. – American Naturalist 127, 358-378.

Mellgren, R. L. and Brown, S. W. 1987. Environmental constraints on optimal-foraging behavior. – In: Quantitative Analyses of Behavior, Volume VI: Foraging. (Commons, M. L., Kacelnik, A. and Shettleworth, S. J., eds) Lawrence Erlbaum Associates, Hillsdale, New Jersey, pp. 133-151.

Mellgren, R. L., Misani, L. And Brown, S. W. 1984. Optimal foraging theory: prey density and travel requirements in *Rattus Norvegicus*. – Journal of Comparative Psychology 98, 142-153.

Murdoch, W. W., and Oaten, A. 1975. Predation and population stability. – Advances in ecological research 9, 2-131.

Oaten, A. 1977. Optimal foraging in patches: a case for stochasticity. – Theoretical Population Biology 12, 263-285.

Olsson, O. and Holmgren, N. M. A. 1998. The survival-rate maximizing policy for Bayesian foragers: wait for good news. – Behavioral Ecology 9, 345-353.

Olsson, O. and Holmgren, N. M. A. 2000. Optimal Bayesian foraging policies and prey population dynamics: some comments on Ridriguez-Girones and Vasquez. – Theoretical Population Biology 57, 369-375.

Pyke, G. H. 1978. Optimal foraging in hummingbirds: testing the marginal value theorem. – American Zoologist 18, 739-752.

Pyke, G. H., Pulliam, H. R. and Charnov, E. L. 1977. Optimal foraging: a selective review of theory and tests. – Quarterly review of biology 52, 137-154.

Riley, D. 2003. A model of optimal foraging: a discrete approximation of continuous time search. – Master's project paper, Department of Mathematics and Statistics, University of Minnesota-Duluth.

Rodriguez-Girones, M. A. and Vasquez, R. A. 1997. Density-dependent patch exploitation and acquisition of environmental information. – Theoretical Population Biology 52, 32-42.

Sernland, E., Olsson, O. and Holmgren, N. M. A. 2003. Does information sharing promote group foraging? – Proceedings of the Royal Society, Biological Sciences 2770, 1137-1141.

Stephens, D. W. and Charnov, E. L. 1982. Optimal foraging: some simple stochastic models. – Behavioral Ecology and Sociobiology 10, 251-263.

Stephens, D. W. and Krebs, J. R. Foraging Theory. – Princeton University Press, Princeton, New Jersey.

Valone, T. J. 1989. Group foraging, public information, and patch estimation. – Oikos 56, 357-363.

Valone, T. J. 1991. Bayesian and prescient assessment: foraging with pre-harvest information. – Animal Behaviour 41, 569-577.

Valone, T. J. and Brown, J. S. 1989. Measuring patch assessment abilities of desert granivores. – Ecology 70, 1800-1810.

Valone, T. J. and Giraldeau, L-A. 1993. Patch estimation by group foragers: what information is used? – Animal Behaviour 45, 721-728.

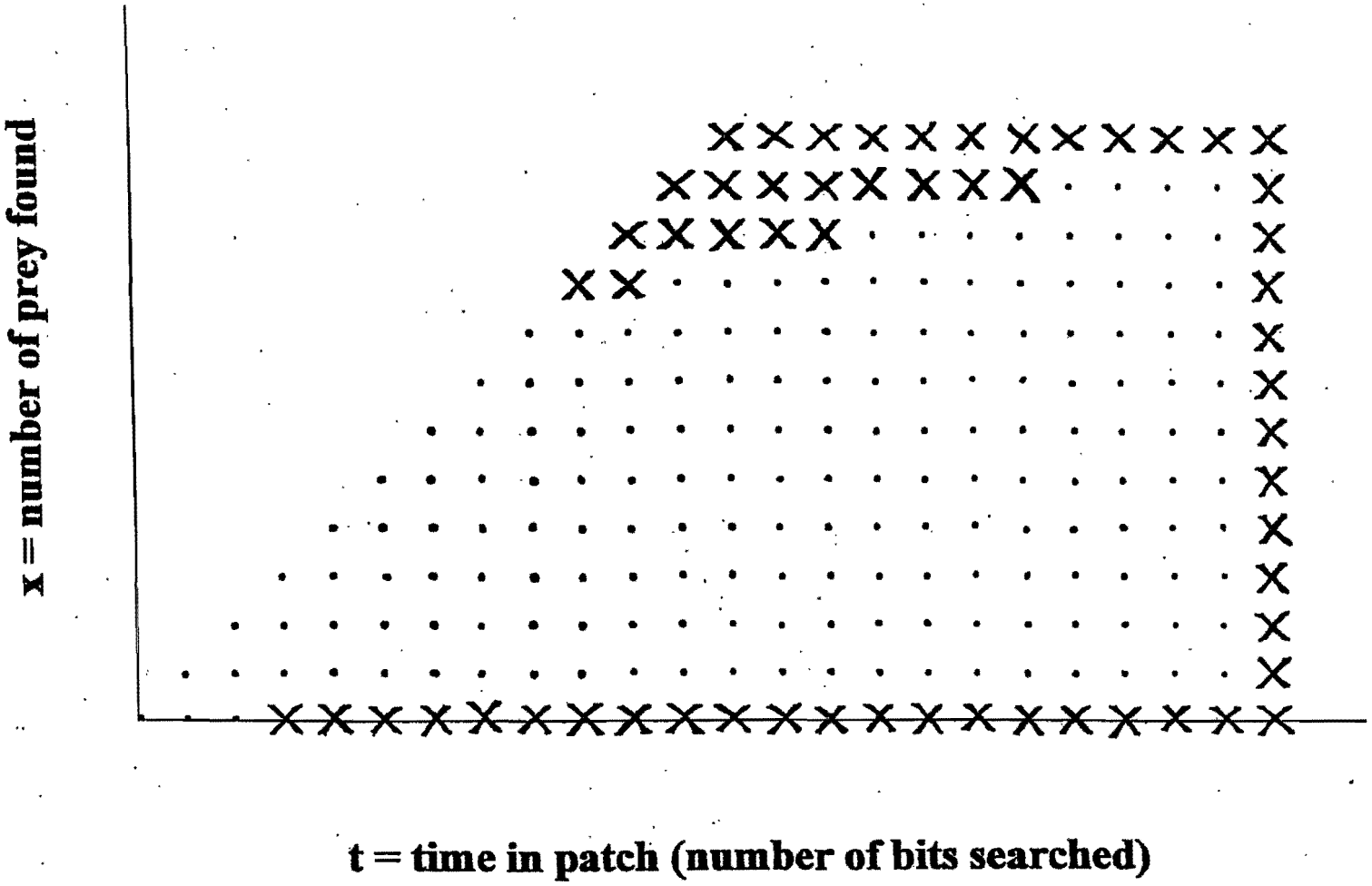
Van Gils, J., Schenk, I. W., Bos, O. and Piersma, T. 2003. Incompletely informed shorebirds that face a digestive constraint maximize net energy gain when exploiting patches. – American Naturalist 161, 777-793.

Wildhaber, M. L., Green, R. F. and Crowder, L. B. 1994. Bluegills continuously update patch giving-up times based on foraging experience. – Animal Behaviour 47, 501-513.

Xu, S. 2000. A model of optimal foraging. – Department of Mathematics and Statistics, University of Minnesota-Duluth, Technical Report 2000-5.

Figure 1

OPTIMAL STRATEGY FOR LIMA'S MODEL



Stopping points (or points after stopping points) are indicated by x

Best rule	**"Lima's rule"**
R = 0.302784	**R = 0.295318**
EG = 4.238364	**EG = 4.278260**
ES = 9.997984	**ES = 10.486956**
τ = 4	**τ = 4**