Lectures on Theoretical Ecology

Simon A. Levin

Department of Ecology and Evolutionary Biology

Princeton University

Stephen P. Ellner

Department of Ecology and Evolutionary Biology

Cornell University

July 25, 2008

Preface

Ecologists try to understand patterns in ecological systems by identifying the underlying processes and mechanisms. Models allow us to make explicit, quantitative predictions about connections between process and pattern, so they provide a framework for developing and testing hypotheses about underlying causes for patterns that we observe. This is not a new idea – one basic result for models of population dynamics is credited to Euler (1707-1783) – but it took a while to catch on in ecology. Beginning in the 1930's, and at an accelerating rate since the 1960's, modeling and quantitative theory have become pervasive in ecology.

The purpose of this book is to help you understand why theory grounded in mathematical models has spread throughout ecology. There is no hope that all of theoretical ecology, or even a significant fraction of it, could be covered in a one-semester course or a single volume of reasonable size. This book represents the lectures that we give when we teach (as each of us does) an introductory course on theoretical ecology. Our syllabi are similar, though not identical, perhaps because both descend from a course that one of us taught and the other TA'd. So this book is more than can be taught to motivated students in one semester, but not much more.

We cover a selection of topics spanning the levels of organization into which ecology is often divided: individual, population, community (interacting populations) and ecosystem. These include some "classical" topics that every theoretical ecologist should know, and some current research. The topics are chosen to illustrate the range of models used by theoretical ecologists, and the kinds of mathematical and computational methods that are used to study them. To introduce core ideas in the simplest possible setting, we often go back to the problems and papers where they originated, and follow that up by jumping straight to recent work. Some bias towards the instructor's research interests and opinions is inevitable, and not necessarily such a bad thing. We teach from the perspective of active researchers. What we therefore cannot give you is the broader and probably more objective perspective of someone observing from the sidelines, or the kind of thorough literature review that we might do if we weren't trying to do research and teaching while also trying to write a book[1]

For the present this "book in progress" is intended to complement course lectures, not to substitute for

---

[1] Robert MacArthur, one of the founders of modern theoretical ecology, used to tell his students (including Mike Rosenzweig, who told us this story) that "You have a choice: you can keep up with the literature, or you can contribute to it". This was in about 1967, and the situation hasn't gotten any better.

them. The contents are still telegraphic in many places, so you need to take your own notes during lectures. Also, don't blame Simon (except for not writing his parts of the book yet); these notes are so far all by Steve. Some sections are based on course notes that were based on the book *Theoretical Evolutionary Ecology* by Michael Bulmer (Sinauer, 1994), which is now out of print.

Chapter 1

Foraging theory

1.1 Testing adaptationism

Darwinian natural selection produces organisms that do a better job at generating offspring, relative to others with whom they compete. According to ecologist Henry Horn (Horn 1971), we should think of plants as "crafty green strategists" who extend branches, grow leaves and sometimes discard them – and so on – in the way that gets them the largest possible share of the light and nutrients that they need for photosynthesis. Herbivores and predators, in this view, are machines for turning "your" offspring into "my" offspring, by eating the former and siring or bearing the latter. The species and genotypes that survive to this day are the ones that have done the best job of this.

The "adaptationist program" attempts to understand organisms' traits in terms of their adaptive value for survival, mating, and reproduction. It looks for the *ultimate* explanation for traits – what purpose are they designed to serve? – rather than the *proximate* means employed to create them, in terms of biochemistry, physiology, and so on. Why do some trees drop their leaves in winter, while others retain them? Why do some organisms typically reproduce several times during their lifespan (iteroparity) while in others reproduction is typically fatal (semelparity)? A hypothetical answer to this kind of "why" question invokes some costs and benefits for the various possibilities, relative to the ultimate goal of leaving the most descendants in future generations.

Models are the means by which such adaptive hypotheses can be tested. To support the hypothesis that trait X is designed to accomplish goal Y, we construct a model to let us predict what trait X *should be* in order to accomplish Y as best as possible. The components of such a model are

1. A *goal function* to be maximized: either Darwinian fitness, a proxy for fitness such as "number of grandchildren that survive to maturity", or some component of fitness such as survival, reproduction, energy intake, etc.

2. The *constraints* under which the organism is operating e.g., what information does it have? what

patterns of movement are possible in its habitat?

3. A set of *options* for the organism.

The goal function has to be computable for any option within the assumed constraints – either by deriving a formula for it, or by constructing a computational algorithm to compute it. Additional assumptions often become necessary at this point, in order to derive a formula or algorithm for the goal function.

Armed with such a model, adaptive hypotheses can then be tested by comparing predictions with observations. This is most often done in a comparative manner:

- By looking across species to see if trait X varies as predicted, as a consequence of differences between species in parameters or constraints, or

- By exposing individuals of one species to a range of conditions and observing how they respond.

Such tests are *qualitative*: we perturb some aspect of the situation, and see if the trait or behavior changes in the predicted direction. So one qualitative test doesn't mean much, but a series of them can add up to strong evidence for or against a theory. Making a quantitative prediction is much more difficult, and rarer.

From the adaptationist perspective, all aspects of ecological systems are shaped by natural selection, so we can seek their ultimate explanation. Our first example will be the structure of feeding relations in an ecological community, which is often summarized by a *food web* such as that shown in Figure 1.1. Like most empirical food webs this one is not completely resolved to the level of species, but it results from what each individual in each species chooses to eat. A community ecologist might take the web as "given" and ask how its structure affects the dynamics of the consituent populations. For example, the wider arrows indicate a loop of interactions (the arrows indicate who eats whom, but the resulting impacts on population growth are two-way), and such loops may have important implications for community stability. But from the adaptationist perspective, we try to understand *why* each species feeds on the ones that it does and not on others; how much time individuals put into feeding versus other activities; how different individuals will organize themselves in space as a result of competition for food; and so on. This is called *foraging theory*, and it is one of the best success stories for the adaptationist program.

1.2   The prey model

The "classical" phase of foraging theory, roughly 1966-1986, centered on a few basic models: the prey (1966), patch (1976), and Ideal Free Distribution (1970) models. The best reference for the classical phase is Stephens and Krebs (1986).

The prey model addresses the choice of which food items a forager should try to capture and eat, so it is often called the "optimal diet" model. As Stephens and Krebs (1986, Chapter 2) observe, essentially
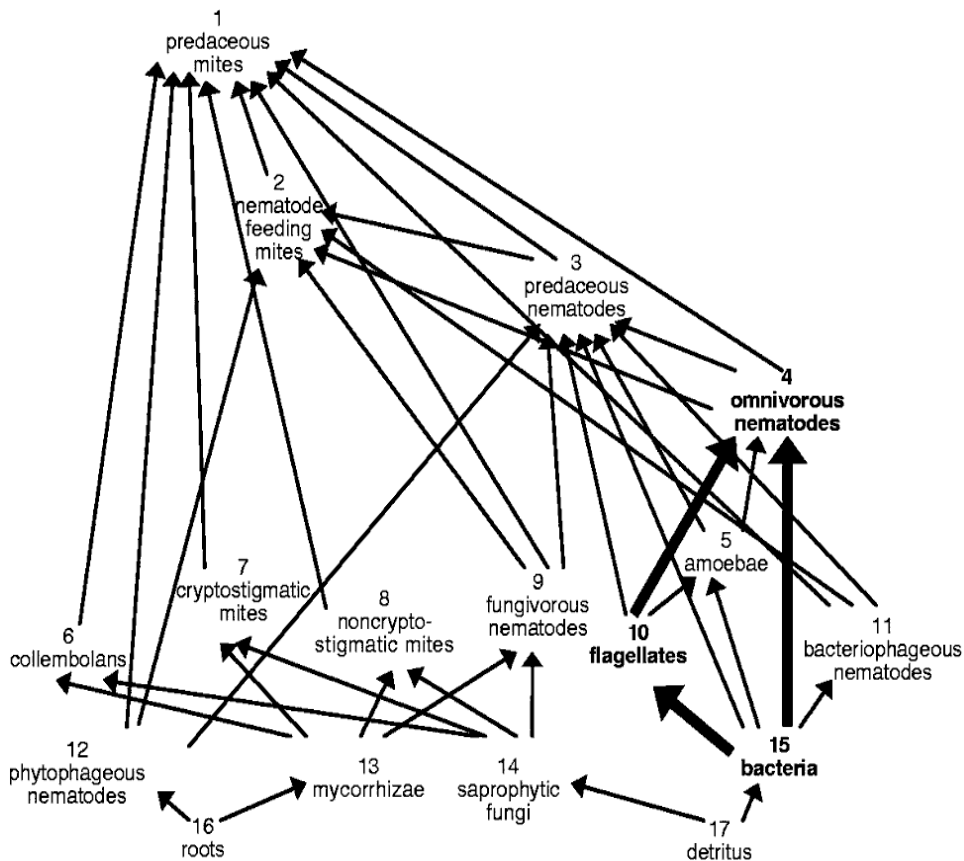
Figure 1.1: Food web summarizing feeding relations in a soil food web (native shortgrass prairie at the Central Plain Experimental Range, Nunn, Colorado) from Neutel et al. 2002.

the same model was proposed independently in half a dozen papers in the early 1970's, building on two similar models proposed independently by MacArthur and Pianka (1966) and Emlen (1966) in back-to-back papers. It was an idea whose time had come.

The model considers a single foraging individual, moving through space in search of food, and encountering prey items of several different types at random. We use "prey" to indicate any potential food item, which includes things like seeds or nectar-bearing flowers.

Model Notation

- i = index for prey type

- λ_i = rate of encounter with prey of type i (items/time = items/area × area/time)

- E_i = net energetic gain from catching and consuming a type-i prey item
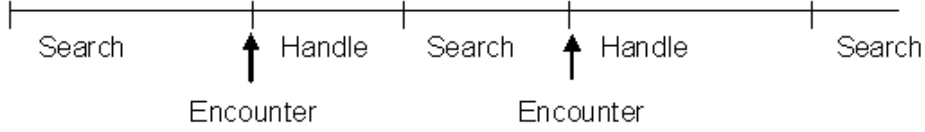
Figure 1.2: The search process assumed in the prey model.

- h_i = handling time for type i items (handle = catch + eat + resume search)

- c = energy cost per unit of time while searching

The classical prey model assumes that the forager knows all of these parameter values, and can identify which type of item it has encountered. If the forager decides to pursue an item, it switches to this activity, i.e. searching and handling are assumed to be exclusive.

Goal: Maximize the long-term rate of energy gain per unit time.

Constraints: The forager can only consume items that it has encountered. If it decides to pursue an item it switches to this activity, i.e. searching stops while the forager is handling an item that it wants to consume.

Options: when an item of type i is encountered, should the forager try to eat it, move on, or "coin toss" and pursue the item with some probability p between 0 and 1?

The forager's "decision variables" are therefore

$$p_i = \text{probability of attacking a type-i item, when one is encountered.}$$

To find the optimal diet, we need to compute the rate of energy gain as a function of the p_i. The forager alternates (by assumption) between searching for prey, and handling, as illustrated in Figure 1.2. S seconds (S large) of searching yields on average

$$S\lambda_i \text{ items of type } i$$

and requires

$$S\lambda_i p_i h_i \text{ seconds of handling time.}$$

Therefore,

$$\text{Total elapsed time is } T = S + S\sum_i \lambda_i p_i h_i$$

$$\text{Total net energy gain is } G = S\sum_i \lambda_i p_i E_i - Sc$$

The rate of energy gain (energy/time) is therefore

$$R = \frac{\sum_i \lambda_i p_i E_i - c}{1 + \sum_i \lambda_i p_i h_i} = \frac{\sum_i \lambda_i p_i E_i^*}{1 + \sum_i \lambda_i p_i h_i} - c \tag{1.1}$$

where $E_i^* = E_i + ch_i$. This informal calculation is rigorously justified by mathematical *renewal theory*. Let each search-and-handle segment of the foraging process be defined as a "cycle". The long term rate of gain is known to be given by the expected gain on a cycle, divided by the expected total time of a cycle. This gives the same result as our heuristic calculation above.

From here on out we drop the *'s on E_j to simplify notation, but remember that E has been adjusted in a way that lets us ignore search costs. We can also set $c = 0$ in equation (1.1), because the value of c has no effect on the optimal choice of p_i. An important principle is involved here: before trying to solve a problem, look for ways to simplify it or make it look simpler. In this case we've found a way to eliminate one parameter, c, by absorbing it into another. So instead of (1.1) we have

$$R = \frac{\sum_i \lambda_i p_i E_i}{1 + \sum_i \lambda_i p_i h_i}. \tag{1.2}$$

Having found R, our goal is to find the p_i that maximize it. We therefore compute $\partial R / \partial p_j$ and find that it has the form

$$\partial R / \partial p_j = [\text{positive}] \times \left(\frac{E_j}{h_j} - \frac{\sum_{i \neq j} \lambda_i p_i E_i}{1 + \sum_{i \neq j} \lambda_i p_i h_i} \right) \tag{1.3}$$

Here [positive] denotes an expression that is always positive, regardless of the values of the p_i. The term in parentheses does not depend on p_j so barring the exceptional situation where it is exactly 0, it is either positive or negative (mathematicians like to consider such exceptional situations, because they often lead to interesting mathematics, but for this class at least we're all biologists). We therefore have

Prediction 1. p_j is either 0 or 1 (no "partial preferences").

The two terms in the parentheses on the right-hand side of (1.3) can be interpreted as the rate of gain from eating a type-j item that has been encountered, and the average rate of gain from moving on with a policy of not eating type-j items. So if it is better to eat the item than move on, $p_j = 1$ is best. Otherwise, $p_j = 0$ is best. Put another way, the optimal diet consists of all prey items j where

$$\frac{E_j}{h_j} > \frac{\sum_{i \neq j} \lambda_i E_i}{1 + \sum_{i \neq j} \lambda_i h_i} \tag{1.4}$$

with the sum running over *all other prey types in the optimal diet*.

Equation (1.4) doesn't actually tell us what the optimal diet is, because you need to know the optimal diet to compute the right-hand side. But it does lead to the guess that foods with a higher value of E_j^* / h_j are better. In fact this is correct:

Prediction 2. Prey types should be ranked by profitability $\rho_j = E_j / h_j$, in the sense that the optimal diet consists of the items with the m highest values of ρ_j for some m.

To verify Prediction 2, we need to show the following: if $E_1 / h_1 > E_2 / h_2$ and type 2 is in the optimal diet, then type 1 must also be in the diet. We use proof by contradiction. So suppose this is not true:

type 2 is in the optimal diet, and type 1 is not. The types in the optimal diet are then $j = 3, 4, \cdots, m$ for some m. Then from equation (1.4) we have

$$\text{type 2 in:} \qquad E_2/h_2 \quad > \quad \frac{\sum\limits_{j=3}^{m} \lambda_j E_j}{(1 + \sum\limits_{j=3}^{m} \lambda_j h_j)} \tag{1.5}$$

$$\text{type 1 out:} \qquad E_1/h_1 \quad < \quad \frac{\sum\limits_{j=2}^{m} \lambda_j E_j}{(1 + \sum\limits_{j=2}^{m} \lambda_j h_j)} = \frac{A}{B} \tag{1.6}$$

where A and B are the numerator and denominator in the last equation. We then have

$$\frac{E_2}{h_2} > \frac{A - \lambda_2 E_2}{B - \lambda_2 h_2}$$
$$E_2 B - \lambda_2 h_2 E_2 > h_2 A - \lambda_2 h_2 E_2 \tag{1.7}$$
$$\frac{E_2}{h_2} > \frac{A}{B}$$

But $\dfrac{E_2}{h_2} < \dfrac{E_1}{h_1} < \dfrac{A}{B}$, a contradiction proving Prediction 2.

So (once and for all) we can label prey types so that $\rho_1 > \rho_2 > \cdots$ and the optimal diet consists of types 1 through m for some m. To find m we can proceed sequentially using (1.4). Start with a "trial" value $\hat{m} = 1$ and ask "should type 2 be added to the diet?". That is, is

$$E_2/h_2 > \lambda_1 E_1/(1 + \lambda_1 h_1)?$$

If no: then m=1.
If yes: then set $\hat{m} = 2$ and ask if type 3 should be added:

$$E_3/h_3 > (\lambda_1 E_1 + \lambda_2 E_2)/(1 + \lambda_1 h_1 + \lambda_2 h_2)?$$

and so on.

This gives us a procedure for finding the optimal diet: it consists of items $1, 2, \cdots, m$ (ranked by profitability) where $m + 1$ is the smallest number such that

$$\frac{E_{m+1}}{h_{m+1}} > \frac{\sum\limits_{i=1}^{m} \lambda_i E_i}{1 + \sum\limits_{i=1}^{m} \lambda_i h_i} \tag{1.8}$$

is **not** true. That is, so long as (1.8) is satisfied by the highest-ranked item among those not already included in the sum, the new potential item should be added to the diet. If (1.8) fails, then the new item (and any of lower rank) should not be added.

Prediction 3. Whether or not a prey type is in the optimal diet depends on its profitability ρ but not its abundance, and on both the profitability and the abundance of more profitable prey items.

Assumptions of the Prey Model

- Search and handling time are exclusive

- Prey are encountered sequentially and at random in proportion to abundance

- Handling and recognition time are constants, unaffected by forager behavior or the abundance of food types in the habitat

- Encounter without attack takes no time or energy

- Perfect knowledge (forager knows the parameters and abundance of all food types)

Predictions

- 0-1 rule (no "partial preferences")

- Ranking by profitability (energy/handling time)

- More selective at higher encounter rates

- Selectivity independent of abundance of low-ranked (non-consumed) items

- Quantitative prediction of threshold profitability for inclusion in the diet

Results

- 51 qualitative tests, 21 quantitative

- 0-1 rule failed in **all** tests.

- Other predictions: 70% agree, 17% disagree, remainder inconclusive

- Mean fraction of assumptions satisfied: 84% for "Agree", 59% for "Disagree"

Table 1.1: **Tests of the Prey (optimal diet) model, based on Stephens and Krebs(1986)**

Prediction 4. An overall increase in abundance of all prey types will eventually increase selectivity (shrink the optimal diet).

Why? Because at each in/out test (1.8), if each of the λ's is increased, the value of the right-hand side becomes larger (verifying this is an Exercise below). That makes each test harder to pass, so m will eventually become smaller.

So does the model work? Stephens and Krebs (1986) reviewed over 70(!) laboratory and field tests of the prey model's predictions, 21 quantitative and 51 qualitative. Here "qualitative" means things like verifying that if you make all prey types more and more abundant, the optimal diet does eventually shrink. The results (Table 1.1) are amazingly good, <u>except</u> for the 0-1 rule.

There is a variant of the prey model in which it is assumed that an item has to be carried back to the

nest in order to be consumed, this is called "central place foraging". The handling time is then increased to $h_i + \tau$ where τ is the travel time back to the nest. Prey are then ranked by $E_j/(h_j + \tau)$, leading to the predictions that:

1. If τ is small, prey types are ranked by E_j/h_j

2. If τ is large, prey types are ranked by E_j

3. If $h_j \equiv h$ then increasing τ shrinks the optimal diet

For these, Stephens and Krebs report that 94% of tests were in agreement with the model predictions.

These analyses were recently updated and extended by Sih and Christensen (2001), who were then able to find 134 published tests of the optimal diet model. They again found very good agreement with predictions (Table 1.2, but identified different factors as determining agreement versus disagreement with the theory:

1. Qualitative versus quantitative tests, with better agreement for quantitative tests.

2. Prey mobility, with better agreement when prey were immobile or partially immobile.

In contrast to Stephens and Krebs, within each type of test (qualitative or quantitative) they found no relationship between the number of assumptions satisfied and how well model predictions were supported. As explanations for these patterns they suggested

1. Quantitative tests require greater knowledge of the study system, reducing the chance of making incorrect predictions due to mis-estimate parameters or some other misunderstanding of the study system.

2. With mobile prey, what a predator actually eats depends on prey behavior – how long and how successfully they can evade capture, for example – so the observed diet is not simply a reflection of the predator's decisions. Within-type variability in the ability of prey to defend themselves when attacked is likely to be large, but is ignored in the optimal diet model.

Predator attack is a strong selective force on the prey, favoring the evolution of defense traits and behaviors that can alter the predator's payoff for each prey type. Even very simple organisms can evolve defense traits – Yoshida et al. (2004) found that traits conferring partial defense against predation evolved in the unicellular green alga *Chlorella vulgaris*, when it was exposed to intense predation by the rotifer *Brachionus calyciflorus*. The impact of prey defense suggested by Sih and Christensen's analysis means that food webs and community structure have to be viewed from a *coevolutionary* perspective: asking what's optimal for the consumer is only half the story.

Exercise 1.1 For the following set of prey items, find the optimal diet according to the prey model.

Factor	Quantitative tests	Qualitative tests
Site of Study		
Laboratory	4.00 (NS)	2.31 (NS)
Field	3.89	2.52
Experiment?		
Yes	3.94 (NS)	2.31 ($P = 0.06$)
No	3.94	2.60
Forager type		
Invertebrate	3.67 (NS)	2.38 (NS)
Ectothermic vertebrate	3.63	2.37
Endothermic vertebrate	4.08	2.47
Prey type		
Immobile	4.22 ($P = 0.08$)	2.61 ($P < 0.005$)
Essentially immobile	4.25	2.40
Slightly mobile	4.00	2.44
Mobile	2.25	2.04

Table 1.2: Effects of different factors on the degree of fit between optimal diet theory and experimental tests, from Table 3 of Sih and Christensen (2001). For quantitative tests the level of agreement was scored from 1 to 6, with a value of 3 indicating qualitative agreement, and 6 indicating perfect quantitative agreement. The range of values for qualitative tests was therefore from 1 to 3. A study was considered to be an experiment if prey abundance was manipulated; all lab studies were classified as experiments. There were 35 quantitative tests and 99 qualitative tests included in these comparisons. Statistical significance of factors was assessed using Mann-Whitney U-test for two-level factors, Kendalls τ for multi-level factors; NS=not significant ($P > 0.05$). Year of study and number of assumptions met were not signficant for either type of test.

Item type	E	h	λ
1	2.6	1.2	4.8
2	2.6	0.7	6.7
3	4.3	1.4	7.3
4	5.3	1.4	7.0
5	7.8	1.9	1.5

Exercise 1.2 Verify the second half of equation 1.1, in which c has been eliminated by defining E_i^*.

Exercise 1.3 Verify that 1.3 is true, by finding the derivative and showing that it can be factored as claimed. If you need to brush up on differential calculus, see the Appendix at the end of this chapter.

Exercise 1.4 Verify that the right-hand side of (1.4) does indeed increase if the abundance of all prey types is increased, for any prey type j in the optimal diet. Hints: (1) the sums in (1.4) only involves prey types in the optimal diet, so we can ignore other prey types. (2) For each prey type k in the optimal diet, note that the right-hand side of (1.4) can be written in the form $(A + B\lambda_k)/(C + D\lambda_k)$. Use calculus to show that this last expression is an increasing function of λ_k for all prey types k in the optimal diet.

1.2.1 Risk-sensitivity

A second failure of the prey model, related to the 0-1 rule, is risk indifference. We assumed that all prey of a given type are identical. But the derivation of R remains the same if E_i and h_i are the average energy value and handling time for prey of type i. Consequently, the model predicts *risk indifference*: the amount of variance in type-specific parameters (with means held constant) should have no effect on the optimal diet.

There is much evidence that foragers are not risk-indifferent. Real (1981) allowed bumblebees and wasps to forage on blue or yellow artificial flowers, in arrays with

- Safe flower type: $2\mu L$ of nectar in each flower

- Risky flower type: either 0 or $6\mu L$ in each flower, with 2/3 of the flowers being empty

The bumblebees exhibited a strong preference for the safe flower type. To explain this Real et al. (1982) postulated that foragers have a nonlinear *utility function* U associated with energy gain. The constant flower yields $U(2)$, the variable one yields $(2U(0) + U(6))/3$. These need not need not be the same.

1. If $U'' < 0$, then $(2U(0) + U(6))/3 < U(2)$ – risk-averse favored

2. If $U'' > 0$, then $(2U(0) + U(6))/3 > U(2)$ – risk-prone favored

If the variance in energy gain X is small, then one can do a "small fluctuations" approximation which says that the long-term rate of energy gain is

$$E[U(X)] \doteq U(\bar{X}) + \frac{\sigma_X^2}{2} U''(\bar{X}) \qquad (1.9)$$

where σ_X^2 is the variance in reward (we'll see later where this comes from). Real et al. (1982) showed that bumblebee behavior was consistent with some predictions from model (1.9). Harder and Real (1987) developed a mechanistic estimate of U for bumblebees, based on things like the time and energy required to extract different volumes of nectar, and found that it worked reasonably well to predict actual preferences for flower types in the lab.

Caraco et al. (1980) found both risk-prone and risk-averse types of behavior in foraging juncos: risk-prone if starved, but risk-averse if well fed before the experiment. This suggests that there is perhaps a *sigmoid* relationship between food reserves and fitness. Caraco et al. (1980) tested this idea by estimating the utility function of well-fed and hungry birds, and found that utility functions were indeed upward-curving for hungry birds, downward-curving for well-fed birds. (Here's one way to determine a utility function by observing behavior, assuming that the function exists and individuals behave so as to maximize their expected utility. Only relative utilities matter so we can take $U(0) = 0, U(6) = 1$ for 0 and 6 seeds obtained at a feeding station. Give birds a choice between a safe station with some constant number of seeds x, and a risky station having either 0 or 6 at frequencies $1 - p, p$. Then adjust p until

the bird is visits the safe and risky stations with equal frequency. This behavior is presumed to indicate that the safe and risky stations have the same utility for the forager. You can then estimate

$$U(x) = (1 - p)U(0) + pU(6) = p.$$

Repeat this for $x = 1, 2, 3, 4, 5$ to fill in the utility function between 0 and 6. If you need to see more of the utility function, increase the reward at the safe station).

Initially much attention was given to simple models like (1.9) that combined the mean and variance of rewards in various ways into an overall reward function. However these have not stood up well against experimental data (Ellner and Real 1989), with one exception: models for "hunting by the hunted" that we will discuss below. The current approach to risk-sensitive foraging is discussed below in section 1.5

Exercise 1.5 For one of the Caraco et al. (1980) birds, the following utilities were estimated for 0 through 8 seeds: (0,0.05, 0.09, 0.18, 0.50, 0.55, 1.00, 1.05, 1.7). Given a choice between two feeder types offering
(1) either 2 or 7 seeds with equal odds
(2) either 1, 4 or 8 seeds with equal odds
which one should it prefer? What number of seeds x would be required for a safe feeder to be preferred over either of these risky feeders?

1.2.2 Predicting food web structure

We introduced foraging theory with the ambitious idea that the structure of real food webs, such as (Figure 1.1), could be predicted on the basis of foraging theory. Beckerman et al. (2006) have suggested that this actually works. Using the optimal diet model and parameters assembled from the literature, they predicted food web connectance $C = L/S^2$ (S=number of species, L=number of trophic links) and how connectance scales with the number of species, for the animal component of a food web. The model leaves out quite a bit – for example, known effects of body size on foraging, and changes in foraging behavior as the density of prey species varies. Nonetheless it did a good job of predicting both the range of C values observed in real food webs, and the relationship between C and S. Connectance is only one aspect of food web structure, but it is an important one for food web stability and robustness against loss of species. This study suggests that empirically parameterized models for adaptive decision-making by individuals may allow us to make explicit, mechanistic connections between individual-level behavioral choices and important ecosystem-level properties.

1.3 Patch model

The patch model, originally due to Charnov (1976), operates on a longer time scale and larger spatial scale. The optimal diet model pretends that the abundance of eacy prey type is constant, even while some of them are getting eaten. An assumption like this is not something we believe to literally be
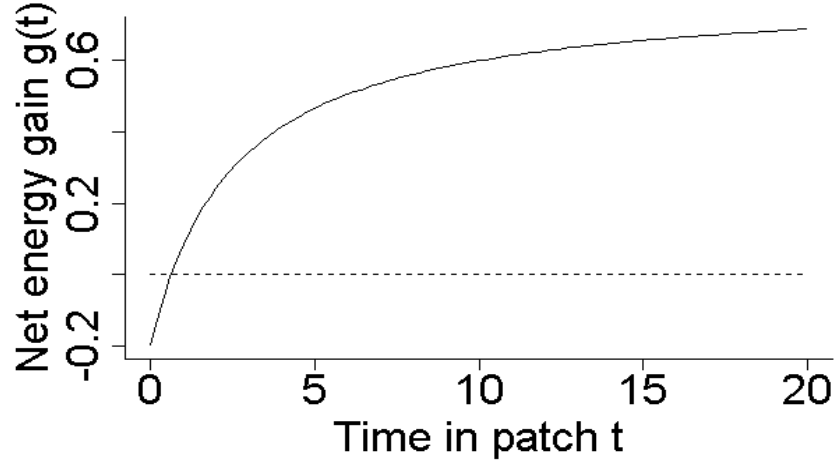
Figure 1.3: The patch model. Note that the energy gain on first arrival in the patch may be negative, to reflect the cost of getting to the patch.

true – it's a simplification that we adopt so that we can focus on one thing at a time. In this case, we focus on a short-term decision whether to attack or not a prey item that is discovered, given the current conditions. But on a longer time scale, a forager (or group of foragers) may deplete the the habitat "patch" where it is now, and move on to another. The patch model addresses the question of when a forager should abandon local search, and try to find a better patch elsewhere.

For now, we assume that all patches are exactly the same. The model is shown in Figure 1.3. The gain function $g(t)$, representing the totl net energy gain after time t in the patch, is assumed to be increasing but decelerating $(g'(t) > 0, g''(t) < 0)$. The only other model parameter is the travel time T required for getting to the next patch.

The forager's decision is how long to stay in the patch, in order to maximize the rate of energy gain

$$R(t) = \frac{g(t)}{t + T}. \tag{1.10}$$

We have

$$\partial R / \partial t = \frac{(T + t)g'(t) - g(t)}{(t + T)^2}$$

so the maximum occurs when

$$g'(t) = \frac{g(t)}{t + T}. \tag{1.11}$$

This condition can be interpreted as follows. The right-hand side is equal to R, the long-term average rate of gain. The optimal "giving up time" is defined by the property that the instantaneous rate of gain within the patch, equals the overall rate of gain in the habitat as a whole. For this reason the solution of
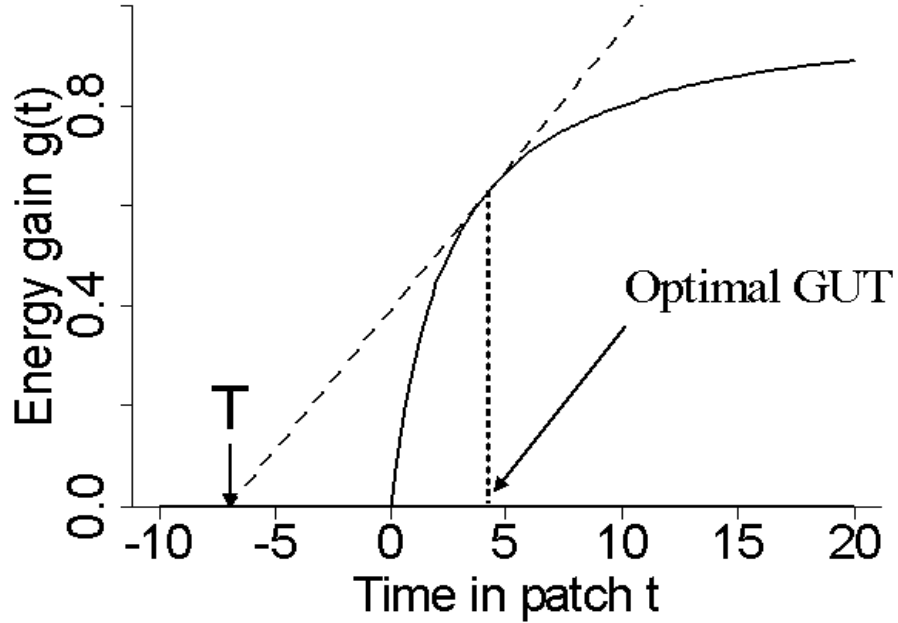
Figure 1.4: Graphical solution of the patch model. "GUT" is the forager's "giving up time", when it gives up on the current patch in order to go find another one

the patch model is called the Marginal Value Theorem (economists use "marginal" to mean "derivative of").

The condition defining the optimal departure time has a graphical solution shown in Figure 1.4. Mark the travel time T as a negative value on the time axis, and starting from that point $(-T, 0)$ swing a straight line down until it is just tangent to the graph of $g(t)$. Let $t*$ be point where the tangent line hits the graph of g. The slope of the line is then $g(t*)/(T + t*)$ (WHY?), which (because it is a tangent line) equals $g'(t*)$.

There is also an extension to multiple patch types. Suppose that the forager can tell what patch type it is in, but only after it has landed in the patch. Let P_i be the fraction of type-i patches and g_i the gain function for patches of type i. Then

$$R = \frac{\sum_i P_i g_i(t_i)}{T + \sum_i P_i(t_i)}$$

where t_i is the giving-up time for patches of type i. As in the prey model, R is given as the expected gain per cycle, divided by the expected time per cycle. With a bit of calculus the condition for optimal giving-up times turns out to be the same in this model as it was for a single patch type:

$$g_i'(t_i) = R.$$

Assumptions of the Patch Model

- Foragers can recognize patch quality (if variable)

- Gain function in each patch type is known and constant

Predictions

- Giving-up time longer when travel time is higher, or mean quality of habitat is lower.

- All patches reduced to the same "marginal value" (rate of energy gain)

- More resources extracted from better patches

- Quantitative comparisons (ala Parker) of observed and predicted giving-up time.

Results

- 24 qualitative tests, 21 quantitative

- 70% agree, 8% disagree, remainder inconclusive.

- Mean fraction of assumptions satisfied=91% for "Agree", 75% for "Disagree"

Table 1.3: Tests of the Patch Model, based on Stephens and Krebs (1986)

Note that this model makes very specific assumptions about what the forager knows about the patch: "nothing" before it lands in it, and "everything" once it gets there. Under different assumptions, e.g. if there is assessment at a distance, things can be very different.

As with the Optimal Diet model, on the whole the Patch Model proved to be quite successful (Table 1.3 summarizes the review by Stephens and Krebs 1986). However, a subsequent review of 26 quantitative tests by Nonacs (2000) found that foragers consistently errer (when there were departures from Patch Model predictions) by staying in patches longer than predicted – cases where they left "too soon" were far rarer. This suggests that simplifying assumptions of the Patch Model lead to a systematic bias in its predictions. Nonacs showed that models incorporating other factors, such as starvation risk and predation risk, predicted longer patch residence times, and he suggested that these factors could explain the Patch Model's bias. Nonacs used Dynamic State Variable models, which are discussed below (section **??**). Carmel and Ben-Haim (2005) developed a very different type of model, based on the assumption that foragers are trying to achieve some required rate of energy gain G_{min}, and behave in the way that minimizes their risk of failure due to uncertainty about parameters. Instead of aiming for optimal and risking catastrophe, their foragers are trying to be good enough with certainty. Their key prediction, like Nonacs (2000), is that patch residence times should be longer than predicted by the Marginal Value Theorem.

The classical test of the patch model was Parker's (1978) study with dungflies, *Scatophaga stercoriara* (Figure 1.3. The "foragers" are males, the "patches" are females, and the "energy gain" is eggs fertilized.

Males congregate at fresh cattle droppings, and attempt to copulate with arriving egg-bearing females. The longer a male remains in copulation, the higher (on average) fraction of her eggs will be fertilized using his sperm; females mate multiple times and an arriving female typically is carrying sperm from previous matings. After mating the male "guards" the female while she deposits eggs fertilized with his sperm, and this is included in the "travel time" (time until he mates with another female). Parker (1978) experimentally estimated the gain curve and the average "travel" (search+guard) time, and found close agreement between the predicted optimal copulation time (42.5 minutes) and the observed average (36.5 minutes).

Parker and Simmons (1994) looked more closely at variation in strategies between males of different sizes. Larger males have a faster rate of sperm displacement (a steeper gain curve) and a shorter search time, because they are able to displace a copulating smaller male and then fertilize most of the remaining eggs. Both of these favor shorter copulation times by larger males. Experimental results confirm this qualitative prediction, and there was also very good quantitative agreement between observed and predicted copula durations, except in the very smallest males. The model predicts very long durations (over 90 minutes) which are not observed. Parker and Simmons (1994) suggest that other factors or constraints might penalize such long durations, such as deterioration of the dung as a substrate for egg laying.

Exercise 1.6 Verify that (1.11) really specifies a maximum of $R(t)$ by computing $\partial^2 R/\partial t^2$, and using (1.11) and the fact that $g''(t) < 0$ to show that $\partial^2 R/\partial t^2 < 0$.

Exercise 1.7 Consider a habitat with patch-types having gain functions $g_i(t) = a_i t_i^{b_i}$, with $a_i > 0, 0 < b_i < 1$.
(a) For a habitat with only one patch type, having $a_1 = 2, b_1 = 0.5$ and travel time $T = 5$, find the optimal giving-up time according to the Marginal Value Theorem.
(b) For a habitat with $T = 3$, 70% patches of type 1 (as above) and the rest of type 2 with $a_2 = 1, b_2 = 0.7$, find the optimal giving up times in each patch.
You could do these by applying the Marginal Value Theorem, but it is more straightforward to do them numerically on the computer (for now, by evaluating the energy gain rate as a function of giving up times over a fine enough grid of values; later on we'll see how to do better).

Exercise 1.8 Best and Bierzychudek (1982) developed a patch-type model for bumblebees foraging for nectar on foxglove. Individual flowers are arrayed vertically on stalks, and (in their experiments) had 10 flowers per stalk labeled 1 to 10 from bottom to top. Bees were assumed to have the following strategy: on arrival to a stalk they checked the bottom flower for nectar. If it is empty they leave the stalk immediately, and if it contains nectar they visit $n \leq 10$ flowers on the stalk (in order from bottom to top) and then move on. The net energy intake rate, as function of n, is calculated as

$$R(n) = \frac{P(\sum\limits_{i=1}^{n} E(i)) - P(n-1)E_w - PnE_f - (1-P)E_e - E_b}{T_b + P(n-1)T_w + PnT_f + (1-P)T_e}$$

where P=fraction nonempty bottom flowers = 0.375
$E(i)$ = mean nectar harvest from ith flower position = 20.19-1.7i cal
$E_b = 0.09$ cal, $T_b = 4.4$ sec are the energy loss and time flying between stalks

$E_w = 0.07$ cal, $T_w = 3.3$ sec are the energy loss and time flying between flowers within a stalk

$E_f = 0.02$ cal, $T_f = 14.7$ sec are the energy loss and time emptying a full flower

$E_e = 0.01$ cal, $T_e = 8.9$ sec are the energy loss and time checking an empty flower

(a) Justify the formula above for $R(n)$. Which of the following assumptions is it making: (1) Flowers on a stalk are independent (some are empty, some are full, with P being the fraction full) (2) Flowers on a stalk are either all empty or all full, and P is the fraction of full stalks.

(b) What is the optimal value of n? (answer this by writing a script file to compute R for any given value of n).

1.4   Ideal Free Distribution

The third basic model is the Ideal Free Distribution (IFD), originally due to Fretwell and Lucas (1970). This model addresses the question of how a group of foragers, each trying to maximize their individual rate of energy gain, should distribute themselves among a group of patches.

Assumptions

1. Foragers are identical.

2. The reward (gain per time) in patch i is $w_i(n_i)$, where n_i is the number of foragers in patch i. This assumption means (tacitly) that resources are constantly being renewed in the patch – each flower makes more nectar, more females dungflies arriving at a dungpat, etc. – and the foragers in the patch are competing with each other for these new resources.

3. w_i is a decreasing function of n_i

4. Foragers are omniscient – in particular they know the reward functions w_i for each patch, and how many others are already in the patch – and are free to move among patches to achieve the highest possible reward, without constraint.

Prediction 1 All competitors experience equal gains, hence the average gain rate is the same in all patches.

To say any more, we need additional assumptions.

Prediction 2 If $w_i = q_i/n_i$ (where q_i is the resource supply rate in patch i) then n_i at steady-state will be proportional to q_i.

This is called the "input matching" or "habitat matching" rule, and has been repeatedly tested. Alternatively, we might suppose that there is interference between foragers. For example, in the prey model we expressed the prey capture rate as (prey/area) $\times$ (area searched)/time. Suppose the area searched per unit time is reduced by the presence of other foragers (giving each other a hard time) say $a_i = Qn_i^{-m}, m > 0$. The prey capture rate in patch i is then

$$\lambda_i = \frac{Qq_i n_i^{-m}}{1 + hQq_i n_i^{-m}}. \tag{1.12}$$

To equalize gains across patches, we therefore must have $q_i n_i^{-m}$ constant across patches, hence $n_i \propto q_i^{1/m}$.

The IFD has had the interesting property that its key prediction - average rate of gain per individual is the same in all patches - has been confirmed even when the model assumptions aren't met, in that there are consistent differences between individuals (Figure 1.6). To account for this, more recent models have incorporated differences between individuals, so that (e.g.) bigger ones make trouble for smaller ones. The trouble with this is that one can make up models *ad infinitum*, and predictions depend on assumptions for which there is still very limited experimental basis. In particular, once you drop the assumption of forager equality, equality of average gain rate across patches is no longer a necessary outcome in the models.

1.5 Dynamic state-variable models

Current foraging theory is largely a response to the failures of the classical models: falsification of the 0-1 rule, sensitivity to risk, effects of variability and interactions among individuals, and the effects of hunger level on behavior .

Mangel and Clark (1986) said "let's bite the bullet" and admit that models including all these complications are too complex for analytic solution. Instead, they proposed using models that can be solved by a computational method called Dynamic Programming. These have the following ingredients

1. Individual state vector $X(t), t = 0, 1, 2, \ldots, T$ (T="final time").

2. The options available to an individual at each time t

3. Dynamic model for changes in $X(t)$ over time, and how these depend on behavioral decisions made at each time step.

4. An <u>additive</u> reward function of the form $W = \sum\limits_{t=0}^{T} g_t(X(t))$

The optimal strategy can be found iteratively using Dynamic Programming. To introduce this method we have to talk about what it was really developed for: trucks (Figure 1.5).

Back to foraging. To see how Dynamic Programming can be used to find optimal state-dependent foraging decisions, we will build a deliberately oversimplified model for a forager having to decide between two patches, differing in food availability (λ_i = probability to find a food item) and the mortality risk (β_i = probability of being eaten). The forager's assumed goal is to remain alive until the final time T. The model components are:

1. State vector $X(t)$ = energy reserves at time t, $0 \leq X(t) \leq 5$.
When a forager falls to reserve level $X = 0$ it is dead of starvation, and $X = 5$ is full: even if more food is found, it can't be eaten.

2. Dynamic Model $X(t + 1) = c(X(t) - 1 + Z(t))$

where $Z(t) = 0$ or 2 is the energy gain at time t, and $c(x)$ is the "chop-off" function

$$c(x) = \begin{cases} 0 & x < 0 \\ x & 0 \leqslant x \leqslant 5 \\ 5 & x > 5 \end{cases}$$

Because $X = 0$ is death, we add the rule that $X(t+1) = 0$ if $X(t) = 0$. The dynamic model says that at each time step a still-living forager uses up one unit of energy. If it finds food in that time step $Z(t) = 2$ and then $X(t + 1) = X(t) + 1$. If it doesn't find food, then $X(t + 1) = X(t) - 1$.

3. Options At each time step the forager can choose to forage in habitat 1 or habitat 2, with parameters $\beta_1 = 0.01, \lambda_1 = 0.5$ or $\beta_2 = 0.05, \lambda_2 = 0.9$. Each of these is "risky" in its own way: Habitat 1 has a high risk of not finding food, Habitat 2 has a high risk of becoming somebody else's food.

4. Fitness (reward) function The assumed goal of surviving to time T means that the reward function is $W(X(T)) = 1$ if $X(T) > 0$, and $W(X(T)) = 0$ if $X(T) = 0$.

We can immediately see that this model leads to *partial preferences* – sometimes the risky patch will be used, and sometimes it won't be. For example, if a forager has $X(T - 1) = 4$ then its best move is to go for the safe patch, since it doesn't need any more food to survive. But if $X(T - 1) = 1$ it should go for the risky one, since the only way it will live is if it gets food, so it's best move is to maximize λ.

To find the optimal decision rule we define

$$J(x, t) = Pr\{\text{survive to time } T \text{ if } X(t) = x\}$$
$$\text{for a forager who behaves optimally between time } t \text{ and time } T. \tag{1.13}$$

$$P_i(x, t) = Pr\{\text{survive to time } T \text{ if } X(t) = x\}$$
$$\text{for a forager who choses patch type } i \text{ at time } t, \text{ and then behaves} \tag{1.14}$$
$$\text{optimally between } t + 1 \text{ and } T.$$

In general, $J(x, t)$ is the expected reward for an individual with state x at time t, assuming optimal behavior thereafter, i.e.

$$J(x, t) = E[W|X(t) = x].$$

J is analogous to "shortest route to the East Coast starting from where you are now" in the truck-routing example. For the reward function defined in item 4. above, the reward is $W = 1$ if the forager survives and $W = 0$ if the forager does not. So the expected reward is the probability that $W = 1$, which is the probability of survival up to time t. In other models, with different reward functions, the definition of J would be modified to correspond to the reward function. Similarly, the functions $P_i(x, t)$ are the expected reward for each of the choices available at time t.

As in the truck-routing example, we find values of J by starting at the end (time T) and moving backwards, alternating between computing values of J and values of P. The process starts by noting

that the value of J at the final time T is determined by reward function W,

$$J(x, T) = \begin{cases} 1 & x > 0 \\ 0 & x = 0 \end{cases}$$

This holds because at time T there are no more decisions to make and no uncertainty as to the final outcome: the forager is either alive ($W = 1\, if\, X(T) > 0$) or dead ($W = 0$ if $X(T) = 0$).

Then how do we move backwards? We know $J(0, t) = 0$ for all t so we only need to consider $x = 1, 2, 3, 4, 5$. We can then compute

$$P_i(x, t) = (1 - \beta_i)[\lambda_i J(c(x + 1), t + 1) + (1 - \lambda_i) J(c(x - 1), t + 1)] \tag{1.15}$$

or in words

$$(\text{not eaten})[(\text{eat}) \times J(\text{new state if eat}) + (\text{not eat}) \times J(\text{new state if not eat})]$$

For a forager with $X(t) = x$, the best decision on day t is to choose the patch type with the larger value of $P_i(x, t)$. We therefore have fitness function

$$J(x, t) = \max(P_1(x, t), P_2(x, t)) \tag{1.16}$$

and optimal patch choice

$$H(x, t) = 1 \text{ if } P_1(x, t) > P_2(x, t), H(x, t) = 2 \text{ if } P_2(x, t) > P_1(x, t). \tag{1.17}$$

So we know $J(x, T)$, and the steps above let us compute $J(x, T - 1)$ for all x. Then we compute $J(x, T - 2), J(x, T - 3)$ and so on, all the way back to $J(x, 0)$.

Typically, as you move backwards in time decisions eventually become a function of state, but not of time: $H(x, T - k) \to \tilde{H}(x)$ when k is large. $\tilde{H}(x)$ is sometimes called the *stationary solution*. In most applications, the model predictions that are reported are based on $\tilde{H}$, a time-invariant rule for how individual decisions depend on individual state.

The initial enthusiasm for these models is best seen in Mangel and Clark (1988), and the outcome in Clark and Mangel (2000). These models proved far harder to test than the classical foraging models: at most a handful of tests are reported in Clark and Mangel (2000), compared to the scores of tests of the basic foraging models in about the same amount of time. The reason for the difference is that there are no general predictions – everything is an inductive generalization from computed optimal strategies in a given model with a given set of parameters. So before you can test predictions you have to make them. That requires constructing the model and estimating the values of its parameters, including the state variable dynamics and fitness function under any possible sequence of decisions. That takes work and time.

For example, Clark and Mangel (2000, Chapter 1) summarize a study of prey choice in sticklebacks feeding on *Asellus* of different sizes. Experiments were conducted to estimate the energy content E_i, handling time h_i, and capture probability for prey of different sizes (3 to 9 *mm* in length), and to

determine the energy requirements, stomach capacity, and energetic search costs of sticklebacks. Prey of intermediate sizes turned out to be the most profitable (largest E_i/h_i), while small prey were easiest to capture. A model was built, similar to our two-patch model, in which the fish's assumed objective was to remain above a critical level of food reserves $X(t)$, with death resulting from falling below the threshold. The model correctly predicted the observed partial preferences for the larger, riskier size classes: well-fed fish concentrated on the smaller prey that could reliably be caught, but fish with low energy reserves and thus a greater risk of starvation would take the gamble of pursuing larger prey when one was encountered.

Another issue is that the power and generality of dynamic programming make it tempting to build a complex model that's hard to understand. Mangel and Clark (1988) built a dynamic programming model based on Charnov and Skinner's (1984,1985) studies of wasps, *Trichogramma* and *Nasonia vitripennis*, encountering hosts for egg-laying of different types (size, species). When a female lays c eggs in a host of type i, the contribution to her lifetime fitness can be measured by the resulting number of *grandchildren* (we count grandchildren rather than number of offspring, because offspring quality (size, vigor) is affected by host-type and egg number).

$$W_i(c) = \# \text{ of grandkids resulting from laying } c \text{ eggs on a type-i host.}$$

Experimental estimates of $W_i(c)$ have a maximum at some intermediate value of c, which is called the "Single Host Maximum" (SHM) because it is the number of eggs to lay if the female encounters only one host of type i. When actual egg-laying behavior is monitored
(1) females lay very variable clutch sizes, and
(2) clutches are almost all below the SHM.
This makes sense. A female should spread her eggs out over multiple hosts to reduce competition between her offspring.

Mangel and Clark (1988) model this by giving each female an initial reserve of R eggs. The model's state variable $X(t)$ is the number of eggs that she has left to lay, so that

$$X(t+1) = X(t) - c(t), X(0) = R. \tag{1.18}$$

The dynamic programming solution is based on $F(x,t)$=expected fitness increment from t to T, for a female with x eggs left to be layed at time t. Let

$$h_i(t) = \text{probability of encountering type-}i\text{ host at time } t. \quad p_t = \text{probability of survival from } t \text{ to } t+1$$
$$\text{(i.e. age-dependent)}$$

So T is defined as the age when p_{T-1} falls to 0. Then the dynamic programming iteration is

$$F^*(x,t) = \sum_i h_i(t) \max_{c \leqslant x}[W_i(c) + p_t F^*(x-c, t+1)] \tag{1.19}$$

and we know that $F^*(x,T) = 0$ (because she's dead).

Experimental data only let Mangel and Clark (1988) estimate W_i, so they generated numerical solutions for a range of plausible scenarios and looked for general properties. They predicted

(1) Older insects (near T) should lay larger clutches (because they are getting closer to the SHM situation).

(2) Due to different host encounter histories, females should have variable clutch sizes.

(3) An increase in mortality favors larger clutches.

(4) As h_i increases, the size of clutches on host-type i goes down.

(5) Females deprived of hosts for a while should then lay larger clutches than ones not deprived, and might use more host-types than a non-deprived female.

The last prediction matches experiments on apple maggot: after deprivation, females more readily used hosts marked by a pheromone showing that another female had already laid eggs there.

However, much of this was achieved with a simpler, analytic model (Parker and Courtney 1984). They assumed:

1. A female has R eggs that she can lay.

2. A female who finds a host has probability p of surviving to find another one.

3. The fitness payoff for laying c eggs on a host is $W(c)$.

From these assumptions, it was possible to derive the expected lifetime fitness as a function of the number of eggs laid per host. This model predicts that

(1) The optimal number of eggs per host is much smaller than the SHM.

(2) An increase in mortality favors larger clutches.

(3) A female who is running out of eggs should lay smaller and smaller clutches.

There will always be a tension between

- Simple broad-brush models that hope to make general predictions by including the most important bits, and hoping that the omitted details won't matter, and

- More complex models that take care to accurately represent case-specific details but therefore make case-specific predictions – generality coming only if general themes emerge from many specific cases.

The argument for the simpler model is that the analytic derivation of predictions lets you see exactly how predictions follow from assumptions. If a prediction is only there because of some dubious assumption that you made for the sake of simplicity, you'll know it and know not to put too much stake in the prediction. This may also lead to new predictions, because you'll see how predictions would differ for other situations where different assumptions would be appropriate. The argument for the more complex model is that you don't have to make so many dubious assumptions for the sake of simplicity, and you still get the conclusions that you need.

But of course you have a third choice – to build both kinds of model and get the benefits of each. Ideally the simpler models can be derived as approximations to the complex model, or as special cases of it,

under some additional assumptions that don't hold exactly but might be close to the truth. The simpler model may then be a guide to the behavior of the more complicated one – providing hypotheses that can be checked computationally – while the more complicated one gives quantitative predictions that can be challenged with experimental data. We benefit from using both kinds of models, and properly used they can be mutually reinforcing.

Exercise 1.9 Write a program that uses Dynamic Programming to numerically solve the two-patch foraging problem with $\beta_1 = 0, \lambda_1 = 0.4$ or $\beta_2 = 0.05, \lambda_2 = 0.8$ for $x = 1 - 5$ from times $t = 99$ to 80, with $T = 100$. ("Solve" means to find, for each x and t, the optimal habitat for the forager to use). At what point (moving backwards from T) does the optimal strategy become time-independent?

Exercise 1.10 . The appeal of dynamic state variable models is ease with which complexities can be added. Patterned on the example above, develop a complete model for a forager deciding between two patch types differing in prey availability λ and mortality risk β), which has both of the following two additional features:

  (i) there is a third option "stay at home" with no chance to get food, or to get eaten ($\lambda_3 = \beta_3 = 0$).

  (ii) the larger an individual is (as measured by their level of food reserves X), the <u>higher</u> their risk of death due to predation while actively foraging (perhaps they can't run as fast, or being larger are more easily detected)

Your description of the model should include a definition of any new variables or functions, all model components (state vector, options, dynamic model, reward function), and the backward-iteration equations analogous to (1.15)-(1.17) that could be used to solve the model and find the optimal strategy.

Exercise 1.11 Suggest another plausible feature that might be added to the basic two-patch foraging model, and predict what changes in optimal individual behavior might result from your modification of the model.

1.6   Hunting by the hunted

Another current aspect of risk-sensitive foraging is "hunting by the hunted". Most animals are at intermediate trophic levels: they are both predators and prey. Going out to hunt for food may place them at greater risk of being detected and eaten. There are a variety of models for different situations, and in some cases simple analytically solvable models have done well. The most successful of these is an approach originated by James Gilliam, based on the hypothesis that foragers minimize their mortality risk subject to getting enough food (to survive, or reach maturity, or whatever they are currently aiming to do).

Gilliam and Fraser (1987) considered a forager choosing among habitats $i = 1, 2, \cdots, n$ with associated energy gain rates $\mathbf{h} = (h_1, h_2, \cdots, h_n)$ and mortality rates $\mu = (\mu_1, \mu_2, \cdots, mu_n)$. The decision is what proportions of time to spend in each habitat, $\mathbf{p} = (p_1, p_2, \cdots, p_n)$. The choice of $\mathbf{p}$ implies

$$\text{Overall mortality rate } U = \langle \mu, \mathbf{p} \rangle$$
$$\text{Overall gain rate } H = \langle \mathbf{h}, \mathbf{p} \rangle$$

The forager's assumed **goal** is to minimize U subject to $H > H'$, where H' is some minimum requirement for energy gain – perhaps the lowest average rate for the day that will allow survival, or allow survival of the forager and its brood of young. The graphical solution is shown in Figure 1.8. The points represent the different habitats. The feasible region (corresponding to different choices of p) is the interior of the polygon (technically, the convex hull of the habitat points). The optimum is the lowest point in the polygon, to the right of the line $H = H'$. This is necessarily on the boundary, so the optimum involves at most two habitats.

Gilliam and Fraser (1987) tested this model experimentally with foraging creek chub minnow, given a choice of 3 habitats: refuge (no food or predators), safe (some food and predators), and risky (more food and more predators). They used the model to predict the amount of food in the risky site that would lead the forager to switch from refuge+safe to refuge+risky as their two habitats of choice, and it worked (Figure 1.9).

Bednekoff (1997) developed a Dynamic Programming model for animals in a group making a decision between foraging and acting as sentinels. Each foragers decision is based on their own food reserve level (similar to the patch selection model) and on the number of others in the group acting as sentinels. A sentinel has a higher probability to detect an incoming predator (in which it alerts the flock and everyone escapes), but a higher risk of being the one attacked if a predator is not detected. A sentinel is performing a "good deed", but in this model its reason for doing so is selfish. The benefit for increased vigilance is assumed to outweigh the increased risk of being in an exposed position, so switching from forager to sentinel decreases an individual's risk of being eaten.

The solution to this model is a "switching curve": a line in the plane with axes $x=$Number of Sentinels and $y=$My Food Reserves (Figure 1.10). An individual above the curve (i.e., one with lots of food reserves) should be a sentinel, and one below the curve should forage. Finding that solution was hard because "my" decision depends on what everyone else's decision rule, so it takes more than just Dynamic Programming – but that's not something we can go into right now. After all this work, Bednekoff's model lets him make the following predictions:

1. Forage when you're hungry. When you aren't hungry, be a sentinel.

2. The higher the predation risk, the more hunger you'll tolerate before you quit being a sentinel and go back to foraging.

It's not clear that a full Dynamic Programming model was needed to reach these conclusions. On the other hand, the paper's predictions were soon put to the test experimentally, and in all cases the predictions were supported (e.g., Clutton-Brock et al. 1999). So there was a real benefit to having a "Swiss Army Knife" like Dynamic Programming that can handle all sorts of decision problems, even if it might sometimes be better to handle them in other ways.

Exercise 1.12 This exercise is inspired by (but not literally based on) Ydenberg and Clark's (1989) study of diving behavior in Western grebes. Imagine a sea bird diving in search of fish, and deciding how deep to go on each dive. The (hypothetical) tradeoffs are

(1) Fish are most prevalent in deeper waters, so the deeper a bird dives, the higher the probability that it will catch a fish on that dive (assume that the bird then has to surface to consume a fish).

(2) A deeper dive entails a higher energetic cost for the dive, and a higher probability of mortality.

Let's say that the energetic gain from catching and consuming a fish is 1 (all fish are equal, and we measure energy in fish units), the probability of catching a fish on a dive to depth y is $y^2/(50^2 + y^2)$, the energetic cost of the dive is $(y/110)^2$, and the risk of mortality is $y/250$ with possible depths $y = 0$ to $y = 100$. For simplicity assume that the time required for a dive is independent of y – not realistic, but this is a homework problem, not a term project. A dive to "depth" $y = 0$ has no cost and no mortality risk – it can be interpreted as a decision to spend a time unit on shore rather than taking a dive.

(a) Build a complete Mangel-Clark style model for state-dependent depth choice during a series of dives at times $t = 0, 1, 2, \cdots, T$ during a day. Make any additional assumptions you need (for example, what is the reward function? – invent something plausible) and define any additional notation needed for your model. As in any such model, you will need to consider a discrete set of possible y values.

(b) Write out and explain the dynamic programming (backwards iteration) equations analogous to (1.15)-(1.17) that could be used to numerically solve your model.

(c) Build, and then find the solution for, a classical-type model that ignores mortality risk and assumes that the bird will dive so as to maximize its long term rate of energy gain.

(d) Build a Gilliam-Fraser style model, which assumes that the forager's goal is to minimize its mortality rate subject to the constraint that the expected rate of energy gain must be at least H. Specifically, give formulas for the expected energy gain and mortality risk as a function of dive depth y and use these to plot the model's graphical representation as in Figure 1.8 (note: you can build your plot by computing the energy gain and mortality risk for $y = 0, 1, 2, \cdots, 100$ and plotting one against the other).

(e) Describe how the value of H affects the optimal strategy in your model from part (d).

1.7   Reminder: the big picture

We have studied foraging theory as a paradigm of the adaptationist program, and it illustrates the structure of any adaptationist theory:

- The assumed goal (sometimes called the "currency", e.g. energy gain per time),

- The assumed set of constraints under which the organism is operating,

- The assumed set of options available (e.g., food types),

- The resulting policy which optimizes the goal subject to the options and constraints.

The last step is where models are essential, and make it possible to test the assumptions by comparison with experiments.

However, "testing optimality models is not an end in itself, but a means to gaining insight into the behaviour of individuals" (Krebs and Kacelnik, Ch. 4 in Krebs and Davies (1991)). We aren't testing whether animals are optimal. We are testing the adequacy of our assumptions about goals, options and constraints – which should include all the ways in which the organism is non-optimal, e.g. not omniscient, imperfect memory, and so on. It's an <u>iterative</u> approach to understanding the "why" of behavioral and other traits as as responses to ecological context, shaped by natural selection.

Even within animal behavior, adaptationist approaches have been applied to many things other than food-gathering:

1. fighting (e.g. males competing for females)

2. mate choice (e.g. males as "prey": mate with this one or search for a better mate?)

3. parental care (stay and raise this brood, or desert to sire another?)

4. conflicts between parents and offspring

5. territory acquisition and defense (how large a territory to defend? how to treat intruders?).

A major focus of research now is on interactions between individuals, in particular on signaling, e.g. chicks begging to their parents for food, and whether showy male traits are reliable as signals of 'quality', and the evolution of cooperative behavior. Many of these require "game theory" models, which we will be studying in a later chapter. There is also an increased attention to the mechanisms whereby goals are achieved, in order to objectively specify options and constraints (i.e. to remove the possibility of getting the right answer by imposing constraints). The latest editions of the books by Krebs and Davies cited below will give a review of current research in adaptationist approaches to animal behavior.

1.8 References

Alonzo SH, P. Switzer P and M. Mangel. 2003. Ecological games in space and time: the distribution and abundance of Antarctic krill and penguins. Ecology 86:1598-1607.

Beckerman, A.P., O.L. Petchey, and P.L. Werren. 2006. Foraging biology predicts food web complexity. PNAS USA 103: 1374513749.

Bulmer, M. (1994) Theoretical Evolutionary Ecology (chapter 6). Sinauer Associates, Sunderland Mass.

Caraco, T., S. Martindale, and T.S. Whitham. 1980. An empirical demonstration of risk-sensitive foraging preferences. Animal Behaviour 28: 820-830.

Caraco, T. 1981. Energy budgets, risk, and foraging preferences in dark-eyed juncos (Junco hymelais). Behavioral Ecology and Sociobiology 8: 213-217.

Carmel, Y. and Y. Ben-Haim. 2005. Info-Gap Robust-Satisficing Model of Foraging Behavior: Do Foragers Optimize or Satisfice? American Naturalist 166: 633-641.

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

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

Clutton-Brock, T.H., M. J. ORiain, P. N. M. Brotherton, D. Gaynor, R. Kansky, A. S. Griffin, and M. Manser. 1999. Selfish sentinels in cooperative mammals. Science 284: 1640-1644.

Ellner, S. and L. A. Real. 1989. Optimal foraging models for stochastic environments: are we missing the point? Comments on Theoretical Biology 1: 129-158.

Emlen, J.M. 1966. The role of time and energy in food preference. American Naturalist 100: 611-617.

Fretwell, S.D. and H.L. Lucas (1970) On territorial behaviour and other factors influencing habitat distribution in birds. Acta Biotheoretica 19: 16-36.

Gilliam, J.F. and D. F. Fraser. 1987. Habitat selection under predation hazard: test of a model with foraging minnows. Ecology 68: 1856-1862.

Harder, L.D. and L.A. Real. 1987. Why are bumble bees risk averse? Ecology 68: 1104-1108.

Heimpel, G.E., J.A. Rosenheim, and M. Mangel. 1996. Egg limitation, host quality, and dynamic behavior by a parasitoid in the field. Ecology 77: 2410-2420.

Horn, H.R. 1971. The Adaptive Geometry of Trees. Princeton University Press, Princeton NJ.

J.R. Krebs and N.B. Davies, An Introduction to Behavioral Ecology. 3rd Edition, Blackwell (1993).

J.R. Krebs and N.B. Davies (eds.) Behavioral Ecology: an Evolutionary Approach. 2nd Edition (1984), 3rd Edition (1991), 4th edition (1997).

MacArthur, R. H. and E. R. Pianka. 1966. On optimal use of a patchy environment. American Naturalist 100: 603-609.

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

Mangel, M. and C.W. Clark. 1988. Dynamic Modeling in Behavioral Ecology. Princeton University Press, Princeton NJ.

McNamara, J.M., and A.I. Houston. 1996. State-dependent life histories. Nature 380: 215-221.

Milinski, M. and G.A. Parker. 1991. Competition for resources. pp. 137-168 in J.R. Krebs and N.B. Davies (eds). Behavioral Ecology: An Evolutionary Approach. Blackwell, Oxford.

Neutel, A. M., J. A. P. Heesterbeek, and P. C. de Ruiter. 2002. Stability in real food webs: Weak links in long loops. Science 296:1120-1123.

Nonacs, P. 2000. State dependent behavior and the Marginal Value Theorem. Behavioral Ecology 12: 7183

Parker, G.A. 1978. Searching for mates. pp. 214-144 in: J.R. Krebs and N.B. Davies (eds.) Behavioural Ecology: an Evolutionary Approach. Blackwell Scientific, Oxford.

Parker, G.A. and L.W. Simmons. 1994. Evolution of phenotypic optima and copula duration in dungflies. Nature 370: 53-56.

Possingham, H. P., A.I. Houston, and J. M. McNamara. 1990. Risk-averse foraging in bees: a comment on the model of Harder and Real. Ecology 1622-1624.

Real, L.A. 1981. Uncertainty and Pollinator-Plant Interactions: The Foraging Behavior of Bees and Wasps on Artificial Flowers. Ecology 62: 20-26.

Real, L., J.R. Ott, and E. Silverfine. 1982. On the trade-off between the mean and variance in foraging: effect of spatial distribution and color preference. Ecology 63: 1617-1623.

Real, L., S. Ellner, and L.D. Harder. 1990. Short-term energy maximization and risk-aversion in bumblebees: a reply to Possingham et al. Ecology 71: 1625-1628.

Sih, A. and B. Cristensen. 2001. Optimal diet theory: when does it work and when and why does it fail? Animal Behaviour 61:379-390.

Stephens, D.W. and J.R. Krebs. 1986. Foraging Theory. Princeton University Press, Princeton NJ.

Verdolin, J.L. 2006. Meta-analysis of foraging and predation risk trade-offs in terrestrial systems. Behav. Ecol. Sociobiol. 60: 457464.

1.9 Appendix: a review of some differential calculus

Differential calculus is a language for working with quantities that change continuously. Here we are concerned with two main examples: the change over time (t) in a variable $n(t)$ describing some aspect of an ecological system, and changes in some system property $M(p)$ as the value of some parameter p (predator death rate, harvesting effort, etc.) is altered.

In general we have some quantity $f(x)$, need to know how much a small change in x impacts the value of f. This is given by the *derivative* of f,

$$df/dx = \text{limiting value as } \varepsilon \to 0 \text{ of } \frac{f(x+\varepsilon) - f(x)}{\varepsilon}. \tag{1.20}$$

The sign of the derivative tells us whether f is increasing $(df/dx > 0)$ or decreasing $(df/dx < 0)$ at x. The numerical value of the derivative tells us how fast this is happening (the "velocity" of f).

Function $f(x)$	Derivative $f'(x)$
x^n	nx^{n-1}
e^x	e^x
$ln(x)$	$1/x$
$sin(x)$	$cos(x)$
$cos(x)$	$-sin(x)$

Table 1.4: Derivatives of some functions.

Other notation for the derivative includes $f'(x)$, $Df(x)$, $D_x f$ and $\dot{f}$ (the last indicating a derivative with respect to time). All of these and others are commonly used.

We also need the *second derivative*, which is the derivative of the derivative – the rate of change of the rate of change, i.e., the acceleration. Notation for second derivatives includes $f''(x)$, $D^2 f(x)$, $D_{xx} f$ and $\ddot{f}$ for second derivatives with respect to time.

Some important properties for us are:

1. At a value of x where $f(x)$ has a minimum or maximum, we have $f'(x) = 0$. If $f'' > 0$ it's a minimum, and if $f'' < 0$ it's a maximum.

2. If $f(x)$ is constant on an interval $a < x < b$ then $f'(x) = 0$ on that interval, and vice-versa.

3. If $f'(x)$ is positive for all x in an interval $[a, b]$ then f is increasing on that interval — if $a \leq x < y \leq b$ then $f(y) > f(x)$. Similarly, if $f'(x)$ is negative for all x in $[a, b]$ then f is decreasing on that interval.

Property (1) is needed in analyzing optimality models (as in: which value of p_j gives a forager the highest rate of energy gain per unit time?). We will use property (2) in analyzing differential equation models, to find values of system variables x at which their rate of change is $f = 0$ – these are then steady states.

For the most part technical assumptions – like specifying whether a function has two continuous derivatives on an interval containing ... – will be omitted. We will always tacitly assume that functions are as smooth as they need to be, random variables have all the moments we need, and so on. To put it another way, we are being more like theoretical physicists than mathematicians.

In order to use these properties of derivatives, we sometimes need to be able to find the derivative of a specific function. Table 1.4 lists a few important simple examples. More complicated cases can often be reduced to these using a few simple rules:

1. If c is a constant, $\frac{d}{dx}(cf(x)) = cdf/dx$

2. **Sum Rule** $\frac{d}{dx}(f(x) + g(x)) = \frac{df}{dx} + \frac{dg}{dx}$ (derivative of sum = sum of derivatives)

3. **Product Rule** $\frac{d}{dx}(f(x)g(x)) = f\frac{dg}{dx} + g\frac{df}{dx}$

4. **Quotient Rule** $\frac{d}{dx}\frac{u}{v} = \frac{v(du/dx) - u(dv/dx)}{v^2}$

5. **Chain Rule** $\frac{d}{dx}(f(g(x))) = \frac{df}{dg}(g(x)) \times \frac{dg}{dx}(x)$

To illustrate the basic rules: we know that $\frac{d}{dx}(x^2) = 2x$ and $\frac{d}{dx}(\sin(x)) = \cos(x)$. From these we get

1. $\frac{d}{dx}(3x^2) = 3 \times 2x = 6x.$

2. $\frac{d}{dx}(x^2 + \sin(x)) = 2x + \cos(x)$

3. $\frac{d}{dx}(x^2 \sin(x)) = x^2 \cos(x) + 2x \sin(x)$

4. $\frac{d}{dx}\frac{\sin(x)}{x^2} = \frac{x^2 \cos(x) - \sin(x) \times 2x}{(x^2)^2} = \frac{x \cos(x) - 2\sin(x)}{x^3}$

The Chain Rule is often written in the simple form

$$\frac{df}{dx} = \frac{df}{dg}\frac{dg}{dx}$$

but you also have to remember where the derivatives on the right-hand side are evaluated. We need the chain rule for functions like $f(x) = \sin(x^2)$. Here $g(x) = x^2$, and $f(g) = \sin(g)$. So to get the derivative we differentiate sin to get cos, and evaluate it at $g = x^2$; then we take the derivative of g with respect to t, which is $2x$. That is,

$$\frac{d}{dx}\sin(x^2) = \cos(x^2) \times 2x = 2x\cos(x^2).$$

It's really not so bad after you've done a few.

1.9.1 Partial derivatives

"Partial derivative" means that we have a function of several variables, and we need to know its rate of change in response to each variable that affects it. How you find partial derivatives is simple: one variable at a time, treating all other variables as constant. A partial derivative is indicated by ∂, for example $\partial f/\partial x$ denotes the partial derivative of f with respect to x

For example consider $F(x,y) = x^2 y$. To take its partial derivative with respect to x, we regard y as a constant. The first of our Rules above then applies:

$$\partial F/\partial x = y \times \partial(x^2)/\partial x = y \times 2x = 2xy$$

Note that for functions of a single variable (such as x^2) the partial derivative is the same as the ordinary derivative. Now what is $\partial F/\partial y$?

We can also take partial second derivatives, which can be "pure" (e.g., $\partial^2 F/\partial x^2$) or "mixed" ($\partial^2 F/\partial x \partial y$). In the latter case, the order doesn't matter: you can first find $\partial F/\partial x$ and then take the derivative of the result with respect to y, or do it in the reverse order (first y, then x).

Of course this new concept brings some new notation. $\partial f/\partial x$ can also be written as $D_x f$ or f_x. For higher partials we have (with the obvious meanings) $D_{xx}f, D_{xy}f$ and so on, or f_{xx}, f_{xy} and so on.

Exercise 1.13 Find the derivatives of the following functions (note that unless otherwise specified log means the base-e "natural" logarithm, and anything else will be specified explicitly, e.g. $\log_{10}$):

$$(1) \quad f(t) = t^3$$
$$(2) \quad g(t) = 2 + t^3 - 5t$$
$$(3) \quad H(x) = x^2 \log(x)$$
$$(4) \quad P(x) = 3x^3 - x^2 \log(x)$$
$$(5) \quad x(t) = \frac{1 + 7t}{1 + t^2}$$
$$(6) \quad f(x) = 2x(1 - x/6)$$

Exercise 1.14 Find the partial derivatives of each function with respect to each of its variables.
(7) $f(x, y) = 1 + 2x^2 + 3x \sin(y)$
(8) $g(x, y) = 1 + ax + by + cx^2 + dxy + ey^2$ where a, b, c, d, e are constants
(9) $P(n, V, T) = nRT/V$, where R is a constant.

1.9.2   Calculus Review Solutions

1. $f'(t) = 3t^2$

2. $g'(t) = 3t^2 - 5$

3. $H'(x) = x^2 \left(\dfrac{1}{x}\right) + 2x \log(x) = x + 2x \log(x)$

4. $P'(x) = 9x^2 - H'(x) = 9x^2 - x - 2x \log(x)$

5. $x'(t) = \dfrac{(1 + t^2)(7) - (1 + 7t)(2t)}{(1 + t^2)^2} = \dfrac{7 - 2t - 7t^2}{1 + 4t^2 + t^4}$

6. Using the product rule $f'(x) = 2(1 - x/6) + 2x(-1/6) = 2 - x/3 - x/3 = 2 - 2x/3$

7. $\dfrac{\partial f}{\partial x} = 4x + 3 \sin(y) \quad \dfrac{\partial f}{\partial y} = 3x \cos(y)$

8. $\dfrac{\partial g}{\partial x} = a + 2cx + dy \quad \dfrac{\partial g}{\partial y} = b + dx + 2ey$

9. $\dfrac{\partial P}{\partial n} = RT/V$ since RT/V is treated as a constant

   $\dfrac{\partial P}{\partial T} = nR/V$ since nR/V is treated as a constant

$$\frac{\partial P}{\partial V} = nRT \times \frac{\partial}{\partial V}\frac{1}{V} = -\frac{nRT}{V^2}$$
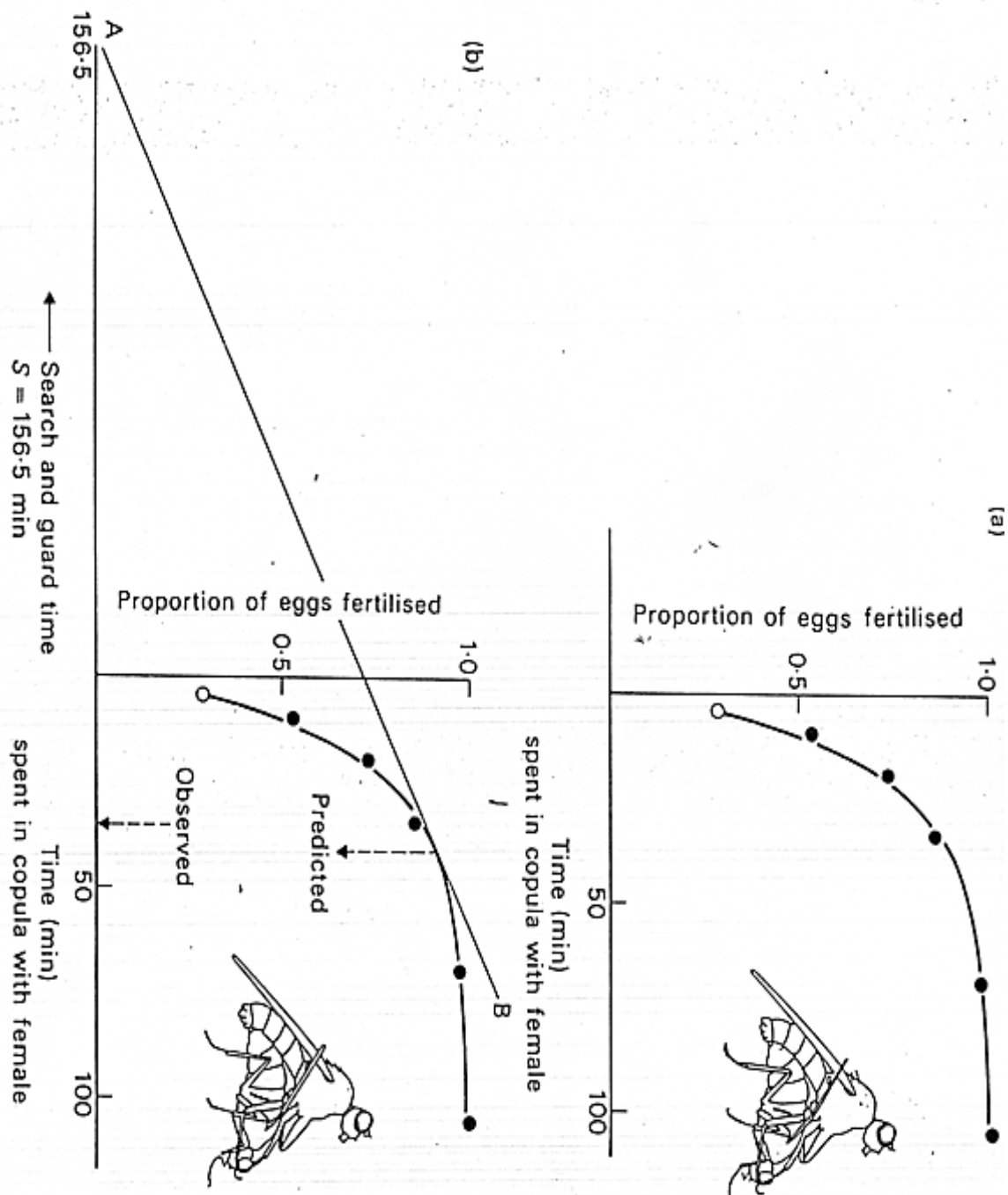
Figure 1.5: Results from Parker's (1978) test of the marginal value theorem using male "dungflies" foraging for egg fertilizations, with females as the "patches".
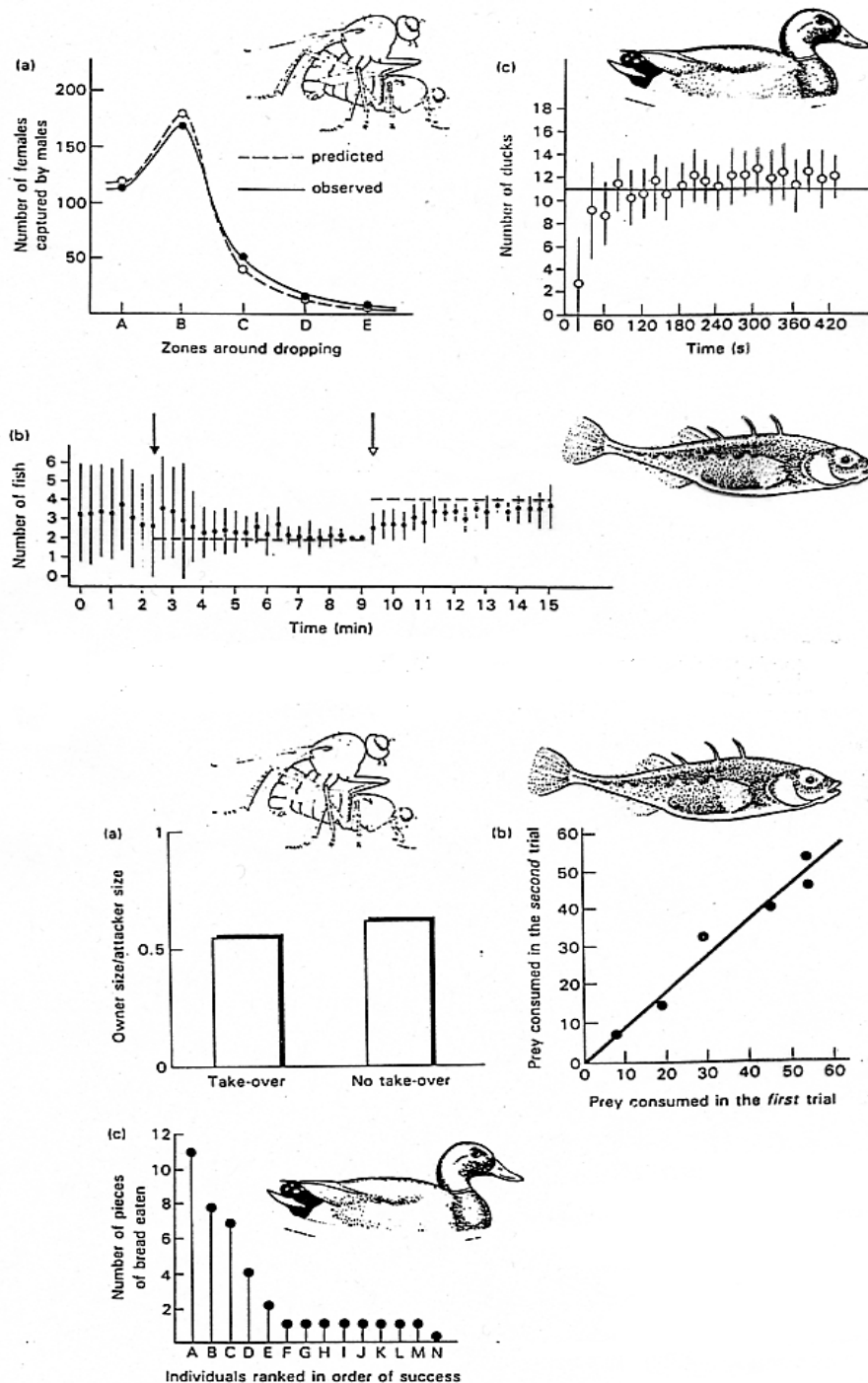
Figure 1.6: Experimental tests of the Ideal Free Distribution, from Milinski and Parker (1991). The upper 3 plots show experiments confirming the key prediction that individuals will distribute themselves so that the mean gain rate is the same in each patch. The lower 3 plots show that none of the studies actually conform to the assumption of forager equality.
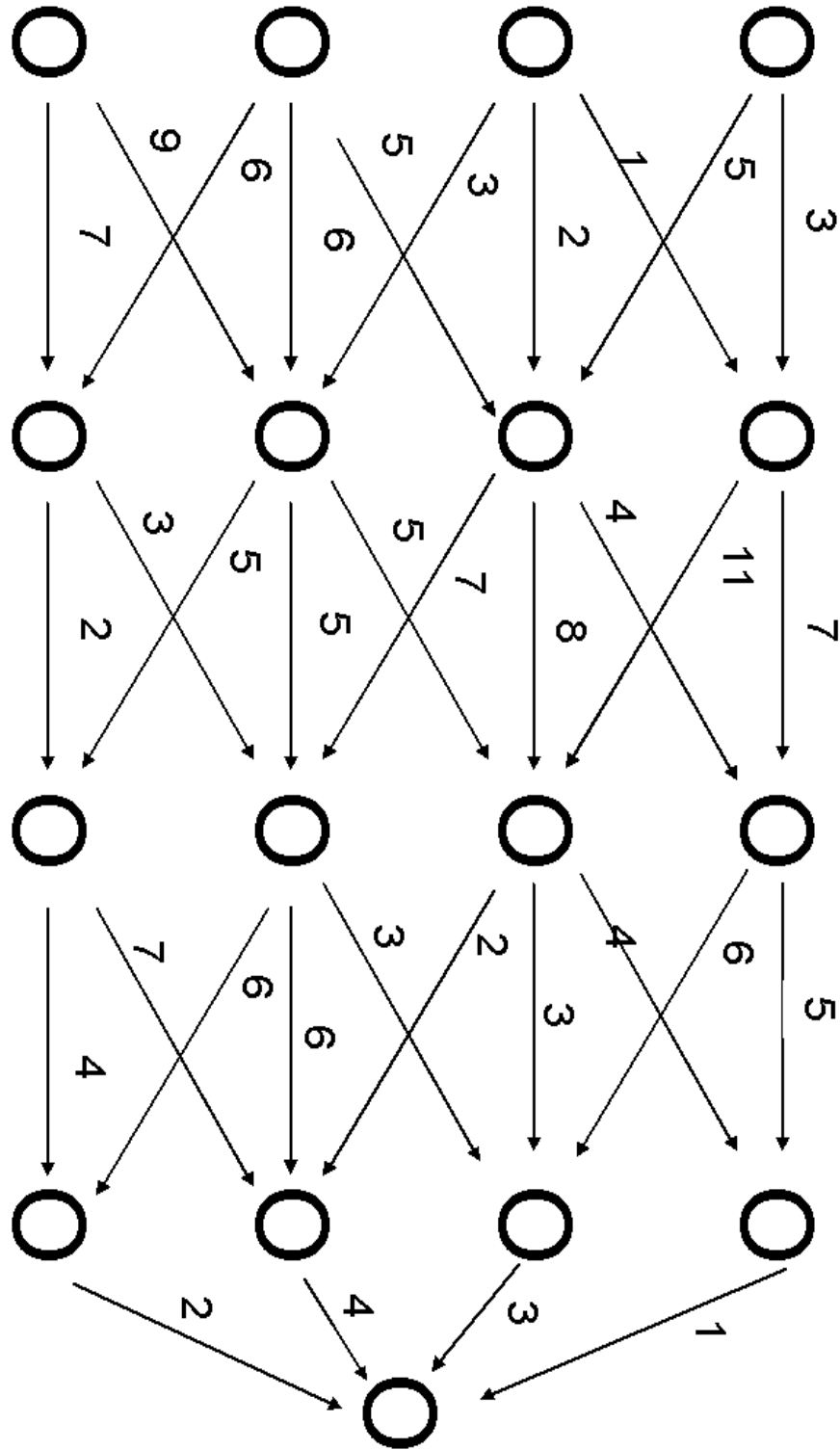
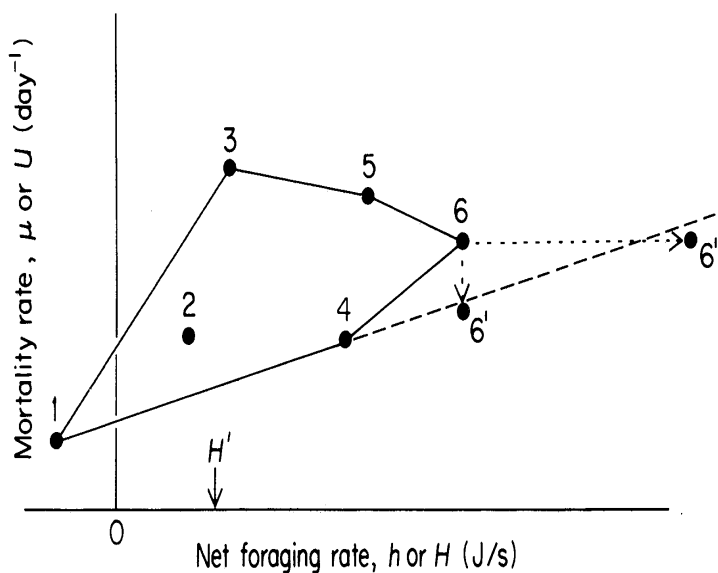Figure 1.7: Problem: find the shortest route from a western node to the eastern node.

Figure 1.8: Graphical representation of the Gilliam and Fraser (1987) model of habitat selection for foraging under predation risk. Each possible habitat is characterized by its net foraging rate and predation risk. Lines connecting habitats represent foragers dividing their time between the two habitats. The optimal strategy is the lowest point of the polygon to the right of the required foraging rate H'. In this figure, the forager should divide its time between habitats 1 and 4. However, a sufficient improvement in habitat 6 – better foraging or lower risk – would lead to the forager abandoning habitat 4 in favor of habitat 6.
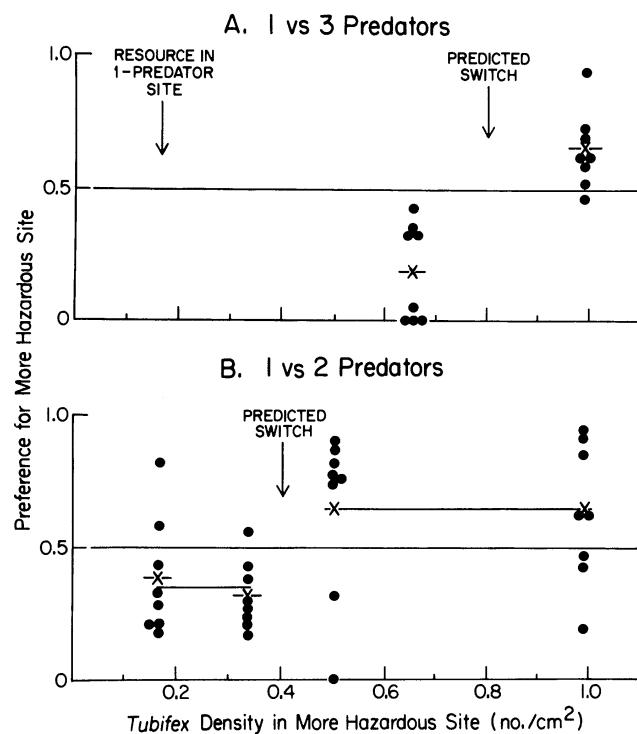
Figure 1.9: Test of the Gilliam and Fraser (1987) model using creek chubs foraging on *Tubifex* worms. The chubs were allowed to choose between 3 habitats: refuge, low-risk, and high-risk (like habitats 1, 4 and 6 in Figure 1.8). As predicted, an increase in the profitability of the high-risk habitat led to a switch from (refuge + low-risk) to (refuge + high-risk).
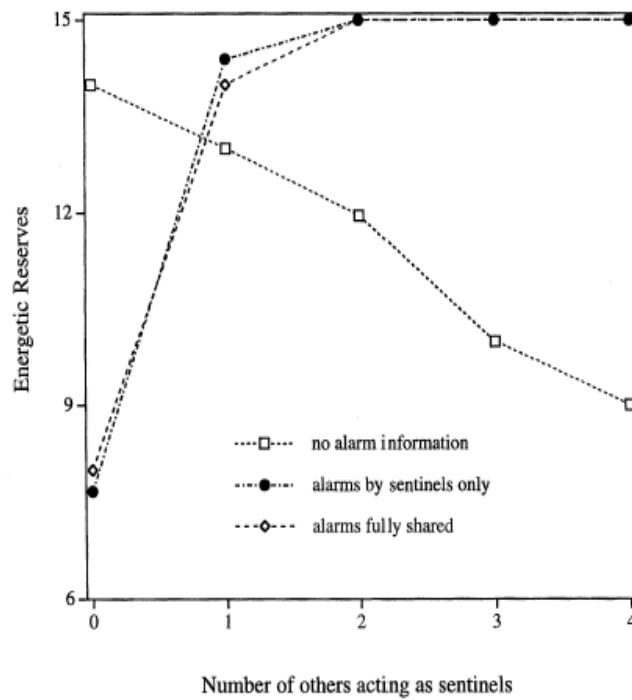
Figure 1.10: A numerical solution of optimal decision rules for the Bednekoff (1997) model of switching between foraging and sentinel behavior. The curves shown are for a group of 5 individuals, with individual food reserves restricted to 15 units or less. Individuals below the curve should forage, those above should be sentinels. The 3 curves correspond to different information about information sharing: no alarm by sentinel or forager who detects a predator (squares), alarms by sentinels only (solid circles) or by sentinels and foragers (diamonds).