

CHAPTER SEVEN

Information: Introduction to Part II.

7.1 Introduction

In this chapter I consider how information is important in foraging theory. My approach is to break the problem into its elements, and to outline and classify these elements and how they are related to each other. In the second section (7.2) I present this outline. I argue that all models of information must treat three components of the problem: what is ambiguous?, how is it ambiguous? and how can the forager gain information to reduce the ambiguity?

In the next four sections, I consider the literature of information in foraging to date, and show how it fits into the framework I have proposed. I discuss the three problems in information which have been studied in some detail, prey recognition costs, patch sampling, and environmental tracking. I also point out a number of problems in information which are completely unstudied.

My treatment is limited to considering fine-grain prey and patch models (that is, models in which encounters, with prey or patches, occur at random according to a Poisson process). For completeness I discuss two problems outside this set of models: habitat sampling, a coarse-grained foraging problem; and models of learning. Finally, I outline the contents of Part II.

The Problem. Reviewers of optimal foraging theory (Pyke et al 1977, Krebs 1978, Werner and Mittelbach 1981) have often concluded that far more work is needed on the problems of information, sampling and learning in foraging theory. However, there has been a surprising amount of work in these areas. Table 7.1 lists and classifies nineteen recent papers related to these problems. Despite this body of work,

TABLE 7.1 OPTIMAL FORAGING PAPERS ON INFORMATION

PATCH SAMPLING:

Oaten 1977	Green 1980	McNamara & Houston 1980
Iwasa et al. 1981	McNamara 1982	

ENVIRONMENTAL TRACKING:

Estabrook & Jespersen 1974	Arnold 1978
Bobisud & Potratz 1976	McNamara & Houston 1980

COSTS OF RECOGNITION:

Elner & Hughes 1978	Hughes 1979	Hughes & Elner 1979
Erichsen et al 1980	Houston et al. 1980	

OPTIMAL MECHANISMS OF LEARNING:

Ollason 1980	Pulliam & Dunford 1980	
Harley 1981	Killeen 1981	Lester in prep.

there seems to be a general consensus that information represents an unsolved problem in foraging theory, (cf. Werner and Mittelbach 1981). The explanation for this is that information can enter foraging theory in many ways; seeking after a single answer to a problem which is best understood as many problems is bound to be a disappointment.

Information and Ambiguity. Although information may affect foraging theory in many ways, the basic plan for all information problems is similar. This plan involves a recognized type of things (patches or prey items) which is itself composed of a number of sub-types which cannot (at least at the moment when the type is recognized) be recognized themselves. I define a "type" to be a recognized class of things, and "sub-types" to be the initially indistinguishable members of this class. I choose the word "type" instead of "class" because it is consistent with the usage of conventional foraging theory, e.g. the patch types of Charnov's (1976) marginal value theorem are patches which can be recognized. A type is said to be ambiguous if it is composed of two or more sub-types. It is assumed that the forager has knowledge of the existence and relative likelihoods of sub-types. In the broadest terms the information problem is: How should a forager treat types it knows to be divided into sub-types? For example, should it pay a cost to recognize them? How much should it pay?

For example, an insectivorous bird may only be able to recognize conifer cones as distinct from other types of food patches in its environment. There are many possible sub-types: fir cones vs. pine cones, or cones with no insects in them vs. cones with ten insects. To make an information model of such a problem, the modeller must provide definite answers to three questions: what is ambiguous? how is it

ambiguous? and how can the forager gain information? At a general level each of these questions can be answered in two ways, giving eight different information problems; two of which are trivial, but of the remaining six only three have been studied at all.

7.2 Elements of the problem.

What is ambiguous? Conventional theory takes as its starting point the moment of encounter with a patch or prey item and asks whether or in what way should this item be exploited. Theory shows that this decision depends on information about two things: the value of the patch or prey item at hand and the value of alternatives. Conventional models have assumed that the forager has perfect knowledge about both. In reality there may be ambiguity about either of these pieces of information.

A model which treats ambiguity in the item at hand asks the question: how should a forager modify its treatment of types when it knows that they are ambiguous. Questions about alternatives are really questions about the environment. Models which deal with ambiguity in alternatives ask the question: how should a forager modify its exploitation of a given item to gain information which may be useful in deciding how to exploit items which have yet to be encountered? At its most extreme, this may mean including types in the diet which are not normally included, to assess whether they should now be included. Ambiguity in the item at hand, and in the alternatives are not always separable questions. For example, there will be conditions in which a forager should modify how it exploits a patch to gain information both about the patch it is in, and about the patches it may encounter in the future. In fact, the extent to which we can consider alternatives and items-at-hand separately depends on the answer to the next question.

How is it ambiguous? When a forager encounters a recognized type it is assumed to know the relative likelihood of each sub-type, which I call the mixing distribution of sub-types. If the forager knows which sub-type it encountered at the i th encounter of a foraging bout, this may give it information about the mixing distributions it will encounter in the future. There may be between-encounter dependencies in the mixing distribution of sub-types, this will usually be because the mixing distribution of sub-types changes in space or in time. On the other hand, it is possible that the mixing distribution of sub-types is the same from one encounter to the next.

There are two qualitatively different types of ambiguity, which I call "encounter-dependent" and "encounter-independent" ambiguity. If the mixing distribution is encounter-independent then I can readily separate ambiguity of alternatives from ambiguity of the item at hand because the item at hand can only give the forager information about itself and not about future encounters. When there are encounter-dependencies the problem of alternatives and items-at-hand cannot be so easily separated. The degree to which they can be separated depends on the problem of interest; for example, it may depend on whether a patch or a prey model is being considered. This problem arises in chapters eight and nine.

How is information gained? There are two mathematically and biologically different ways in which a forager may gain information about which sub-type is which. Firstly, foragers might be able to distinguish sub-types without making any decision about whether or how to exploit the patch or prey item at hand. In other words, foragers may be able to use cues (possibly at some cost) to recognize sub-types, in the way that the shore crabs (Carcinus maenus) studied by Elner and

Hughes (1978) can distinguish profitable from unprofitable sub-types of mussels by "hefting" them. Secondly, a forager may gain information about sub-types while exploiting them. In other words, the forager may sample. Sampling may, or may not, allow a forager to have perfect recognition of sub-types, but it is different from prior recognition. The difference is due to when the price of recognition is paid. If the price of recognition can be paid before the item (patch or prey) is exploited, then we can logically separate the costs of recognition from the gains due to exploitation. In sampling, the costs of recognition are closely tied to the way the item is exploited. The cost of recognition and the value of exploitation cannot be readily separated. This interaction occurs in two ways. Firstly, sampling depletes (although sometimes insignificantly) the value of the patch or prey item sampled. Secondly, the degree and cost of recognition due to sampling is strongly dependent on the nature of the sub-types and the mixing distribution of sub-types.

For example, imagine that a particular patch type is composed of three sub-types. Each sub-type has at most one prey per patch, the sub-types differ only in the rate (λ) at which the prey can be found $\lambda = 0, 1, 2$ prey/min, according to a negative exponential distribution. The three sub-types are equally abundant. After .93 minutes without finding a prey the forager will be 90 percent certain that the patch is not of sub-type $\lambda = 2$. It will have to search another one and a quarter minutes to be 90 percent certain that there are no prey in the patch. The forager has to pay quite different prices (measured in searching time) for the same amount of certainty about each alternative.

In table 7.2 I present this scheme of classification as a branching diagram. The table shows that there are eight possible ways

TABLE 7.2 A CLASSIFICATION OF INFORMATION PROBLEMS

What is Ambiguous?	How is it Ambiguous?	How is Information Gained?	Work to Date
Item-at-Hand	Encounter Independent	Recognition	<u>I</u> Hughes 1979 Elner & Hughes 1978, Houston et al. 1980, Erichsen et al. 1980.
		Sampling	<u>II</u> Oaten 1977 Green 1980, Iwasa et al 1981, McNamara 1982, Chapter ten.
	Encounter Dependent	Recognition	<u>III</u> Unstudied.
		Sampling	<u>IV</u> Unstudied.
Alternatives	Encounter Independent	Recognition	<u>V</u> Trivial Problem
		Sampling	<u>VI</u> Trivial Problem
	Encounter Dependent	Recognition	<u>VII</u> Unstudied.
		Sampling	<u>VIII</u> Estabrook & Jespersen 1974 Arnold 1976 Chapters eight & nine.

to answer the three questions posed above. In the following sections I consider how the information literature to date is related to these eight cases of the information problem. I divide the problems which have been studied into three: the cost of prey recognition, patch sampling, and environmental tracking.

7.3 The value of recognition.

A number of papers (Hughes 1979, Elner and Hughes 1978, Erichsen et al. 1980, Houston et al 1980) have considered modifications to conventional diet models in which the forager faces encounter-independent ambiguity, and is assumed to be able to recognize sub-types perfectly by paying a price in the currency of recognition time, in cases of (category I of table 7.2). The theoretical parts of these papers illustrate that in cases when recognition can be paid for, the required changes to conventional diet theory are straightforward, but by no means trivial. Perhaps most importantly, Elner and Hughes, working with shore crabs (Carcinus maenus); and Houston et al and Erichsen et al. working with great tits (Parus major) have studied the costs of recognition experimentally.

Cost of recognition problems like these represent the most tractable of the six non-trivial cases of the information problem, as illustrated in table 7.2. Moreover, the experimental work cited above shows that there are natural situations in which this simplest of information problems may arise.

Mathematically, it is possible to ask questions about recognition of encounter-independent ambiguity at a more general level. Specifically, one can ask the question, how much should a forager be willing to pay for perfect recognition of sub-types? Gould (1974) has studied this problem in the context of human decision makers in

microeconomics, by defining the "value of information." I prefer to call Gould's formulation the "value of prior recognition," when it is applied to foraging theory. I now outline Gould's analysis, and discuss two of its implications.

Gould's value of prior recognition. Gould (1974) imagines that the world is comprised of several indistinguishable sub-types, $(s_1, s_2, \dots, s_k)$, and that an individual, a firm, or in our case a forager is making some decision (i.e. choosing a decision variable Y) about how to behave, for which the pay-off is determined by which sub-type actually obtains. That is, the pay-off is some function of s_i and Y , $H(s_i, Y)$. The decision-maker knows the relative likelihood of each sub-type $(p_1, p_2, \dots, p_k)$, i.e. it knows the mixing distribution of sub-types. If the forager has no further information it should choose $Y=Y^*$, where

$$\sum_{i=1}^k p_i H(s_i, Y^*) = \max_Y \sum_{i=1}^k p_i H(s_i, Y) \quad (7.1)$$

Y^* is the optimal value of Y . However, if the decision maker recognizes that the sub-type it is faced with is a particular sub-type s_i before choosing Y , it should choose $Y=Y_i^*$, where

$$H(s_i, Y_i^*) = \max_Y H(s_i, Y) \quad (7.2)$$

These two expressions allow Gould to answer the question: Given that the decision-maker knows only the relative likelihoods of sub-types (the p_i 's), how much should it be willing to pay (in the currency of H) to have perfect recognition? Gould's value of information is the difference

$$\sum_{i=1}^k p_i H(s_i, Y_i^*) - \sum_{i=1}^k p_i H(s_i, Y^*) \quad (7.3)$$

However, in foraging it is better to call Gould's formula the value of prior recognition; it could, for example, be used to generalize cost-of-recognition models, (e.g. Hughes 1979). Gould's definition allows him to make several important and counter-intuitive points about information.

Gould considers how the mixing distribution of sub-types affects the value of recognition. If all the p_i 's but one are set to zero, there is only one sub-type, and the value of recognition is nil. However, what mixing distribution maximizes the value of recognition? Gould shows, against initial expectations, that the value of recognition is maximized where there are exactly two sub-types. (Gould's "two sub-type" result is not as puzzling as it seems. For example, imagine that there are only three conceivable patch types: 0 prey, 5 prey or 100 prey per patch. Imagine that there are twelve different habitats, and that the forager knows that each habitat has a different permutation of patch types. Gould's results says that the forager should be willing to pay most for perfect recognition in one of the six habitats where only a pair of sub-types exists. In this case almost certainly the '0 prey per patch-100 prey per patch' habitat.)

Gould's value of prior recognition is defined in terms of a particular decision problem. This is in contrast to other measures of information (e.g. DeGroot 1970, Shannon and Weaver 1949 both cited in Gould 1974) which consider only the decrease in environmental ambiguity due to information gain. The Shannon-Weaver index commits what might be called the academic fallacy: it assumes that certainty is intrinsically valuable. However, in foraging theory (and similar decision problems) certainty is only valuable if it increases the forager's rate of energy intake (or other currency of maximization).

For our purposes, it is important to deduce from Gould's definition that for a forager recognition is of limited value; the suggestion is not that recognition is unimportant, but that there may be some cases in which foragers should not bother with recognition. This is a conclusion which I expect to be broadly applicable in information problems in optimal foraging theory. Information must be valued in terms of the decision problem at hand, and this value will, in general, be limited.

7.4 Patch sampling.

Imagine, as above, that there is ambiguity in the item just encountered by a forager, and that this ambiguity is encounter-independent. But instead of being able to pay a price for prior recognition of sub-types, the forager can only gain information by actually beginning to exploit the item at hand, (II in the classification of table 7.2). This is the conventional patch sampling problem. The problem of recognition applies equally to patches and to prey items. However, sampling the item at hand applies most directly to patches; because it is usually assumed that prey items can only be completely ingested or ignored (there is no intermediate level of exploitation of prey items). Although ingesting a prey item may reduce ambiguity about the sub-type at hand, as long as the ambiguity is encounter-independent there is no point in sampling prey items (this problem re-appears in chapter ten). On the other hand, a large patch may be used in many ways, ranging from spending only a few seconds to many minutes in the patch. In this case, it may be valuable to use information gained in the first moments of patch exploitation to make better use of the patch.

The patch sampling problem is the most studied theoretical problem

of information in optimal foraging. Specifically, this problem has been studied using a variety of models by Oaten (1977), Green (1980), McNamara and Houston (1980), Iwasa et al (1981) and McNamara (in press). All of the work to date has been theoretical, and has largely centred on the extent to which patch sampling can be used to refute Charnov's (1976) "perfect patch recognition" model. This issue is confused by the impression left by Oaten (1977; cf. Werner and Mittelbach 1981) that this is in some sense the stochastic problem in optimal patch use. As I have pointed out, it is only one problem out of six in information, and information is only half of the stochastic problem in general, (see Stephens and Charnov in press for a discussion).

Of the models of patch sampling, the most elegant and general is McNamara's (in press) discussion of the "potential" value of information. McNamara illustrates his general model with a special case in which the only difference between sub-types is the initial number of prey per patch. In this case information is gained at each prey capture. McNamara shows that a forager should leave a patch when:

$$r(t) + p(t)h(t) = R \quad (7.4)$$

where t is the time spent so far in the patch, $r(t)$ is a stochastic version of Charnov's marginal capture rate, $p(t)$ is the subjective probability that a prey item will be found at time t , and $h(t)$ represents the potential increase in the forager's overall rate of gain due to the information gained if a prey item is found at time t ; R is the average rate of gain in the environment. Notice that if the product $p(t)h(t)$ is zero then this is a stochastic version of Charnov's "marginal-value" theorem. $h(t)$ would be zero if the patch sub-type is perfectly recognized; because finding a prey item could not change the

forager's assessment of the patch. On the other hand, $h(t)$ might be large if the mixing distribution of patch sub-types has high variance; for example, if patches have either one or ten prey in them, the second prey item has a greater information potential than if the only sub-types are one-prey or two-prey patches. In this case, we can think of the term $p(t)h(t)$ as determining the degree to which patch sampling modifies the forager's optimal strategy, (the comparison isn't perfect because $r(t)$ also contains information about the patch).

Perhaps the most significant result of the patch sampling literature is that patch sampling will be most important (compared to non-sampling strategies) when the variance in the mixing distribution of patch sub-types is high, as suggested above. I discuss this idea in detail in section 10.3. One difficulty with these sampling vs. non-sampling comparisons is that the models compared make different assumptions so that it is not clear whether the comparison is fair. For example, Green (1980) criticizes Charnov's (1976) model of optimal use of "depressing patches" (this is the terminology of Charnov et al 1976) by using a model without patch depression. In chapter ten I apply an analysis similar to Gould's to see whether these changes can be made independently of the value of patch sampling.

7.5 Tracking the environment by sampling.

Suppose that, unlike the two cases studied above, a forager faces some type of encounter-dependent ambiguity, and that information can only be acquired via sampling. This is, in outline, the problem of tracking a changing environment by sampling, (VIII of table 7.2). In this case information gained at one encounter may affect the forager's assessment of the next encounter, that is, its assessment of the value

of alternatives. Put another way, conventional foraging models usually assume that the forager knows some parameters of its foraging environment (for example, the over-all rate of energy gain); how should a forager track encounter-dependent ambiguity in these parameters?

This argument has lead Pyke et al (1977) to call this problem "parameter estimation," suggesting that a straightforward application of the statistical theory of parameter estimation may yield a solution to the problem. This view is mistaken for two reasons. Firstly, as with Gould's results, the value of knowing a parameter must be defined in terms of a given foraging problem. A statistician wants to know the value of μ ; a forager only cares if knowing μ increases its rate of energy gain, (which it will not always do). Secondly, statisticians conventionally assume that they are sampling from an unknown, but unchanging probability distribution. In environmental tracking problems, if the probability distributions do not change then there is nothing to track. For these reasons, the general environmental tracking problem is in the more esoteric reaches of statistical theory.

A better line of attack is to consider the optimal use of information in conditions of encounter dependent ambiguity. Two models of such problems have been studied. McNamara and Houston (1980) have studied a single switch problem, in which the forager is assumed to know that a given type will switch from giving a unit reward at a fixed probability to giving no rewards. McNamara and Houston use this model to predict how long a run of bad luck the forager should "put up with" before switching to some alternative food source. They show that their predictions are supported empirically by the operant phenomenon called the partial reinforcement effect, (Mackintosh 1974). Such single switch problems are the simplest and most analytical environmental

tracking problems, and it seems surprising that they are so little studied experimentally and theoretically.

A more general model of sampling a changing environment has been proposed by Estabrook and Jespersen (1974), and studied by Bobisud and Potratz (1976) and Arnold (1978). This model predicts the optimal sampling period (by analogy to physics, period is the inverse of sampling frequency; i.e. "sampling period" is the interval between sampling events) in a changing environment. This model allows for the possibility of multiple switches by imagining a very simple rule for changing sub-types from one encounter to the next. This model differs from McNamara and Houston's in another important way; it assumes that the forager can recognize a change of sub-type perfectly after a single experience with the type. These two models may be thought of as two halves of a more general problem: How should the forager sample to recognize a change? (McNamara and Houston) and How should it sample when a change is easily recognized? (Estabrook and Jespersen).

In chapter eight I analyse the Estabrook and Jespersen model of optimal sampling period theoretically. I present a simpler and more intuitive solution to Estabrook and Jespersen's, than those previously presented. In chapter nine, I present the results of an experiment designed to test the Estabrook and Jespersen model.

7.6 Some unstudied information problems.

Encounter dependencies. To some extent the distinction between information about alternatives, and information about the items at hand breaks down in the previous section. This is because there will be situations in an environmental tracking model when information from previous encounters will give information about the item at hand, as well as about the alternatives. Perhaps I should have classified the

problem of the previous section as categories VIII and IV of table 7.2. However, the separation of category IV is justified for two reasons. Firstly, in an environmental tracking problem knowledge of previous encounters may provide information about the type at hand, but always provides information about alternatives. Secondly, category IV describes a more subtle and unstudied problem: how should patches be exploited in light of information from previously encountered patches. Biologically, this might be called the "patch of patches" problem. In a patch model, encounter dependencies may mean that the forager is in a good or bad patch of patches. If the forager has just exploited a number of extremely good patches this may affect its behaviour at the next patch encountered. This problem is unstudied. Notice that like in the tracking problem patches may have an "alternatives" information value as well as an "item at hand" information value. Past encounters may modify the way the forager should exploit the item at hand, because they give it information both about alternatives and this item.

Prior recognition and cues. A quick glance at table 7.2 suggests that information gained by recognition is under-studied. This may be because, compared to sampling problems, it seems trivial. It is a much more tractable problem than sampling, but as the work of Gould (1974), Hughes (1979) and Houston et al (1980) shows it is not trivial when recognition comes at some cost.

At a more general level the use of information gathered by means other than direct sampling is virtually unstudied. How should foragers use information gained from the environmental correlates of their prey items? This is important, because such correlates are abundant, and probably comparatively cheap to use. Kushlan (1978) has shown that wading birds are attentive to each other's fishing success;

nectar feeders could easily use the time of day to decide which flowers to exploit, since nectar depletion follows a predictable daily pattern, (e.g. Pyke 1978a).

Alternatives without encounter dependencies. In table 7.2, I have marked categories V and VI "trivial problems." Imagine that the forager could perfectly recognize all sub-types in the environment, (that is, perfectly recognize alternatives). If all the environmental ambiguity is encounter-independent, then all this recognition can give the forager is a list of the relative frequencies of sub-types, (specifically, under encounter-independence the forager's assessment of what sub-type will be encountered next cannot be changed). As long as there are enough items in the environment to satisfy the law of large numbers this list of relative frequencies will be the same as the mixing distribution of sub-types, which the forager already has. Therefore, the forager gains nothing by this perfect knowledge.

7.7 Habitat sampling and rules for learning.

Habitat sampling. In the previous section I have considered information in the two problems of fine-grain foraging theory, patch-use and diet choice. Information may be important in coarse-grain problems as well. The use of a simple patch or prey item occupies only a small fraction of the forager's time budget, before it goes on to another. A forager may also have to decide which of several possible habitats it should exploit, where a habitat is defined as a "patch" so large that an individual's foraging causes only insignificant depletion. For example, bumblebees may need to decide which of two meadows to exploit, as discussed by Heinrich (1976).

Since there is no depletion of habitats a rate maximizer only needs to know the rate of energy gain associated with each habitat

type. If these rates are known perfectly the forager should exploit the habitat type with the highest rate, so that in the perfect knowledge case the habitat selection problem is trivial. If there is some ambiguity about habitats then the forager may need to sample to find which is best. Krebs et al (1978) have studied this problem experimentally and theoretically, using a case where the forager faces a choice of two habitats. They ask the question (in rather a limited way), how long should the forager sample before deciding to exploit one habitat or the other? Two parameters are shown to be theoretically important, the difference between the unknown reward rates, which determines how readily the habitat sub-types can be distinguished; and the time the forager has available for foraging, (the "time horizon," Krebs et al. 1978). Krebs et al.'s experimental results confirm the importance of these variables, although there is still room for theoretical and experimental refinement, (cf. Kacelnik 1979, and Houston et al in press).

Rules for learning. Many behavioural ecologists seem to equate the information problem with learning. There are two difficulties with this. Firstly, learning is a confusing term because it is used in two ways in foraging theory. There are models of "learning how," in the sense that a bumblebee may learn to exploit a complicated flower (Heinrich 1976); one implication of "learning how" for foraging theory has been discussed by Hughes (1978). There are also models of "learning about" in the sense that a forager may "learn about" which sub-type it is dealing with; this is more or less the information problem as discussed in this chapter.

A second problem with the term "learning" is that the conventional approach of learning theorists has been mechanistic, or causal (see

Mackintosh 1974). Learning theorists have searched for general rules which govern the learning process, or how information gained from experience is used to modify behaviour, as opposed to how information should be used as I have discussed above. Many models have approached the causal question by assuming that each behaviour in a mutually exclusive list of behaviours is played according to a probability, and that it is this probability that is updated by experience, (for review see Krebs et al in press). Such models seem reasonable only in the search for general mechanisms for the use of experience to change behaviour. There is no reason to believe that these models represent a "best" solution to this problem until more is known about the costs of using different types of learning mechanisms. In fact, there are reasons for holding these models at arms length. For example, the information problem as discussed above suggests that the search for a general rule may be futile, since equivalent experiences (say finding a prey item) may require extremely different changes in behaviour (it may mean the forager should leave immediately in one patch type, and stay in another). Another difficulty is that when behaviours are only played according to a probabilistic rule, the nature of patterned behaviour which may be produced is limited. For example, a fixed sampling period, which may be a very sensible strategy (cf. chapter eight), cannot be produced.

7.8 Summary, Discussion and Outline of Part II.

In this chapter I present a scheme for classifying the elements of the information problem in fine-grain models of optimal foraging. Specifically, I propose that three questions must be answered in any model of information. 1. What is ambiguous? There may be ambiguity either in the item (patch or prey item) at hand or in the

alternatives. 2. How is it ambiguous? Ambiguity may be encounter-independent or encounter-dependent. 3. How is information gained? Foragers may gain information either by paying a price for prior recognition or by sampling.

I present this scheme as an aid to delineating the elements of a complex problem, rather than as a hard and fast taxonomy. If it is useful, future workers will almost certainly find that it requires modifications and refinements. For example, I have pointed out that weakness of the distinction between "alternatives" and "items-at-hand" when there is encounter-dependent ambiguity.

This scheme may be most useful to point to deficiencies in the state of the art. For example, as general problems, encounter dependent ambiguity and cost of recognition (as opposed to sampling) models seem little studied, although they both arise in important and natural ways. Outside of this scheme, I discuss the the coarse-grain problem of habitat sampling, and models learning mechanisms.

Outline of Part II. In the second part of this thesis I discuss two problems. In chapters eight and nine I re-consider the environmental tracking model of Estabrook and Jespersen (1974), and in chapter ten I consider a number of aspects of the patch sampling problem.

In chapter eight I present a fairly simple argument to clarify the problem of environmental tracking, specifically the choice of an optimal sampling period, and I re-formulate the Estabrook and Jespersen model in a special case. In chapter nine, I present an experimental test of this model of optimal sampling period using great tits (Parus major) foraging in an aviary.

In chapter ten I consider the comparison between sampling and non-sampling models of patch-use by an analysis similar to Gould's analysis of the value of prior recognition. I define the value of patch sampling, and I use this definition to study three problems: the effect of variance in the mixing distribution of patch sub-types; the importance of alternative and recognizable patch types; and the relationship between patch depression and patch sampling. Finally, in chapter ten I develop a model of partial patch recognition, in which some aspects of the patch can be recognized upon encounter but others cannot.

CHAPTER EIGHT

Tracking a Changing Environment: Some Simple Theory

8.1 Introduction A forager may need to track some attribute of its environment which changes with time. When a forager encounters a prey item it needs information about the value of the item encountered, and about the value of the alternatives. If the value of alternatives changes with time then (even if they should be ignored for the present) it may be wise for the forager to check the alternatives periodically to see if it should change its mind about excluding them. In this chapter, I ask the question, how frequently should a forager check alternatives? The surprising result of this analysis is that there is a remarkably small range of conditions in which the forager should bother with checking alternatives at all.

Elements of the tracking problem: Imagine that a forager is sampling a fluctuating system. Consider a prey type which fluctuates between two sub-types, good and bad, according to some probability model. Both sub-types are defined by probability distributions of prey value. If these distributions overlap then sampling may not perfectly distinguish good from bad. Distinguishing sub-types is the first element of the environmental tracking problem. Secondly, imagine that by some sampling scheme the forager can perfectly distinguish good from bad (this will not always be possible). The forager still must decide whether to exploit this varying prey type, or to ignore it for the present, and if it is ignored, then when should the forager check it again? One might say the forager must still decide on optimal treatment of the varying type.

These two elements of the problem are related in an interesting way. If sub-types cannot be distinguished at all the problem of optimal

treatment is trivial; the varying type should be treated as if it were an unambiguous type according to its mean profitability. Similarly, sampling the system can only be reasonable if the cost of distinguishing the sub-type (via sampling) is sufficiently small. Broadly speaking, the more difficult (or costly) it is to distinguish sub-types the less important sampling will be as a component of optimal treatment.

In this chapter I consider optimal treatment of a varying prey type, making very simple assumptions about the forager's ability to distinguish sub-types by sampling. I assume that sub-types can be distinguished perfectly after only one sample. It is important to bear in mind that this simplification is in favour of sampling behaviour, so that the case considered is extreme. I present this model in two sections. In section 8.2, I consider the optimal sampling period (period $\approx 1/\text{sampling frequency}$) for a forager encountering a simple kind of immediately distinguishable run of bad luck. This model is chiefly of interest as a simple illustration of the important features of the environmental tracking problem. In section 8.3 I review a more general model developed by Estabrook and Jespersen (1974) and Arnold (1978), and studied by Bobisud and Potratz (1976), I consider a more restrictive case than that treated by previous authors. The model substantiates the simpler analysis of section 8.2. In chapter nine, I present an experimental test of the Estabrook and Jespersen model.

8.2 Optimal treatment of a run of bad luck.

Imagine that foraging consists of a sequence of encounters with a prey type called X, (so that time is measured in encounters with X). The forager does not have to eat X's because there are always alternative prey available. Normally, type X is better than the

alternatives, providing food of value v_g (the value of good sub-types) at each encounter. Rarely, X 's change in value to v_b (the value of bad sub-types). In terms of chapter seven, type X is composed of two sub-types "good" and "bad." The alternative prey have value v_a per encounter. Assume that $v_g \geq v_a \geq v_b$. When X switches from good to bad the length of the run of bad luck (L), measured in encounters, is a geometrically distributed random variable, $P(L=i) = (1-q)q^i$. L is the number of bad encounters after the first, so $i = 0, 1, 2, \dots$ etc. q is called the stationary probability of the system. The run of bad luck is a first-order Markov process, that is, the probability that X is bad at encounter $i+1$ depends only on the state of X at encounter i . The stationary probability $q = P(X_{i+1} \text{ is bad} \mid X_i \text{ was bad})$; similarly, I call $1-q = p = P(X_{i+1} \text{ is good} \mid X_i \text{ was bad})$ the transition probability. The Markov property has special implications; observing two bads in a row leaves the forager with no more information than observing one (cf. Bobisud and Potratz 1976). Similarly, every bad observed gives the forager the same information about the expected run length of bads. This means that the best way to treat the X -type has an especially simple form. After each bad observed the forager should ignore X 's for exactly N encounters, where N is the sampling period. The problem is to find the optimal sampling period N^* .

What factors should influence the choice of N^* ? Figure 8.1 shows a run of bad luck and the behaviour of two foragers, one adopting $N=4$, the other adopting $N=12$. An omniscient forager would simply ignore X 's for L , the run length, encounters. Compared to this omniscient behaviour any forager adopting a sampling period N , must make some errors. The crucial thing is to notice that there are two types of

Figure 8.1
The Trade-off between the costs sampling and sensitivity to change.
Finding an optimal sampling period.

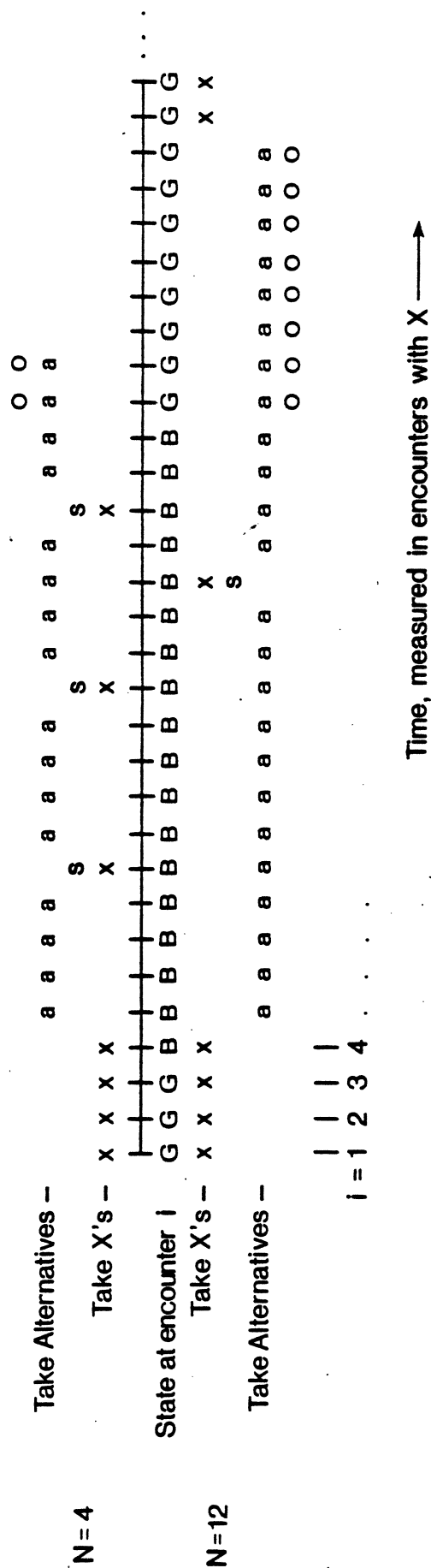


Figure 8.1 The figure shows a sequence (solid bar) of good (G) and bad (B) sub-types: a run of bad luck. The forager deals with this by ignoring the type for a fixed number (N) of encounters every time a bad (B) is observed. N=4 in the upper case; N=12 in the lower case. There are two types of errors: sampling errors marked S and over-run errors marked O. When N is small there are few S's and few O's. When N is large there are few S's and many O's.

errors associated with an N-sampling period. Firstly, when a forager samples X 's and finds they are still bad, it gets v_b units of food when it could have had v_a , the value of alternatives. I call this a sampling error (marked S in figure 8.1). Sampling errors cost $v_a - v_b$. Secondly, the forager may ignore X 's for too long, in which case it makes an error in getting only v_a units of food when it could have had v_g . Those over-run errors (marked O in figure 8.1) cost the forager $v_g - v_a$ units. The figure shows that small sampling periods ($N=4$ in the figure) are quite responsive to the change back to good, but produce many sampling errors. Conversely, large sampling periods ($N=12$ in the figure) produce few sampling errors but many over-run errors. The optimal sampling period N should reflect this trade-off between the cost of sampling errors (sampling too often) and the cost of over-run errors (not being responsive enough to change).

The analysis of this section is based on minimizing the total expected cost of these errors. I show that the ratio of the cost of sampling errors to the cost of over-run errors is the crucial variable in determining the optimal sampling behaviour. I call this ratio b and define it as,

$$b = \frac{v_a - v_b}{v_g - v_a} \quad (8.1)$$

This b is logically equivalent to the b used by Estabrook and Jespersen(1974) and Arnold (1978). However, this definition is unique in taking explicit account of the value of alternatives (see section 8.4 for discussion).

Let N_s be the random variable which describes the number of sampling errors, and let N_o describe the number of over-run errors. It is readily seen (from figure 8.1) that N_o bears a systematic

relationship to N_s and L of the form

$$N_0 = (N + 1)(N_s + 1) - L - 1 \quad (8.2a)$$

Since this equation is linear in the two random variables, N_s and L ,

$$E(N_0) = (N + 1)\{E(N_s) + 1\} - E(L) - 1 \quad (8.2b)$$

I wish to minimize, by choosing the sampling period N , the expected cost of mistakes. I do this by minimizing the equivalent term

$$M_N = bE(N_s | N) + E(N_0 | N) \quad (8.3a)$$

or by substitution.

$$M_N = E(N_s | N)(b + N + 1) + N + E(L) \quad (8.3b)$$

A sampling error occurs only if the run of bad luck lasts for the first $N + 1$ encounters; similarly, there are i sampling errors, if and only if, the first i groups of $N + 1$ encounters contain no switches back to good and the $(i+1)^{st}$ group of $N + 1$ encounters does contain a switch. This means that the number of sampling errors, N_s , is also a geometrically distributed random variable with stationary probability q^{N+1} . Therefore

$$E(L) = \frac{q}{1 - q} \quad (8.4)$$

$$\text{and} \quad E(N_s | N) = \frac{q^{N+1}}{1 - q^{N+1}}$$

Since the sampling period is a discrete variable the analysis must proceed by difference equations. M_N is decreasing with N if

$$M_N - M_{N-1} < 0 \quad (8.5a)$$

by substitution and some algebra I find that this is equivalent to

$$b > \frac{1 - q^N}{q^N(1 - q)} - N = n(N, q), \quad (8.5b)$$

where N must be greater than zero, and where $n(N, q)$ denotes the right

hand side of the inequality. It is possible to show that $n(N,q)$ always increases with N

$$n(N,q) - n(N-1,q) > 0 \quad (8.6a)$$

$$\text{or, } \frac{1 - q^N}{q^N(1 - q)} - \frac{1 - q^{N-1}}{q^{N-1}(1 - q)} > 0 \quad (8.6b)$$

which is algebraically equivalent to

$$(1 - q)(1 - q^N) > 0 \quad (8.6c)$$

which must always be true if $q < 1$. The procedure for finding the optimal sampling period is straightforward. Given conditions of q and b , N^* is the optimal sampling period if

$$n(N^*,q) \leq b < n(N^*+1,q) \quad (8.7)$$

N^* is an increasing step function of b . For a given q , N^* can be found with a hand calculator using $n(N,q)$. Figure 8.2 shows the optimal sampling period, N^* , as a function of the error ratio b for two values of the stationary probability q , ($q = 0.95$ and 0.75).

As general rules we may deduce that as b increases, N^* will increase, and that increases in q will also increase N^* . Both results seem reasonable and expected on the basis of the trade-off between sampling and over-run errors. The importance of this model is to point out the central role played by b in models of environmental tracking. Although this model allows us to see this clearly; it is also mistaken in at least one assumption. I have assumed that runs of bad luck are extremely rare, so that I could treat them as isolated events. I have consequently ignored the possibility that X 's will change to good and back to bad again while the forager is ignoring X 's. The model explored in the next section allows for this possibility.

8.3 The Estabrook-Jespersion-Arnold model

In this section I analyse a special case of the more general model

Figure 8.2 Optimal Sampling Period as a Function of the Error Ratio.

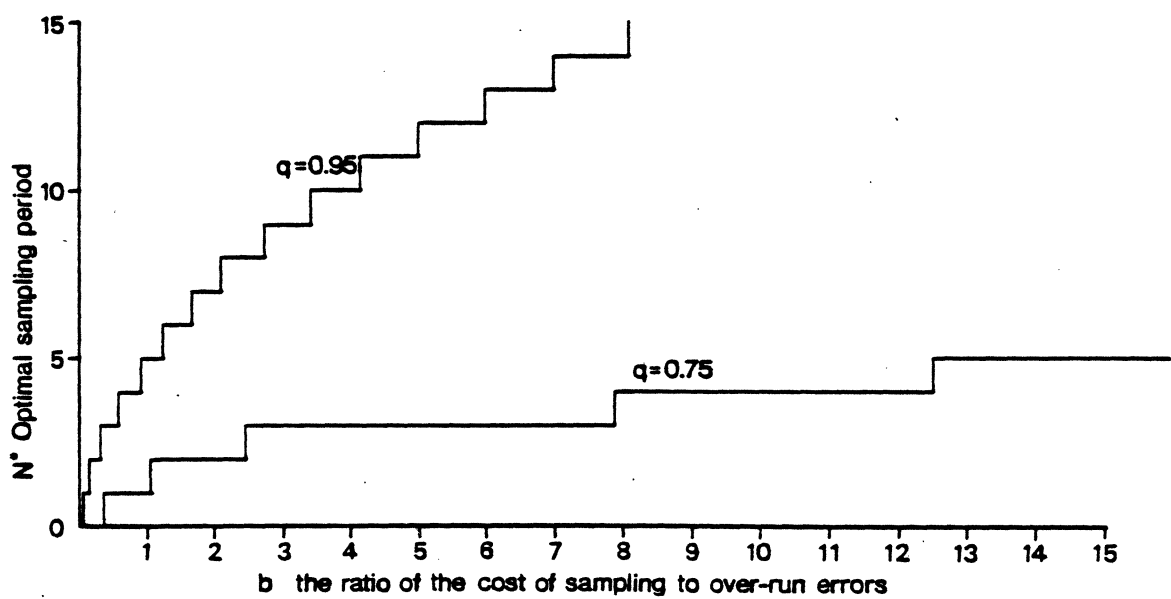


Figure 8.2 The figure shows the optimal sampling period N^* as a function of b the ratio of the cost of sampling errors to over-run errors; derived from the simple run of bad luck model of section 8.2

of optimal sampling period first proposed by Estabrook and Jespersen (1974). This model is broadly similar to the model in the previous section. Estabrook and Jespersen were interested in the degree of protection from predators a model-mimic system could achieve if the predator had a flexible sampling rule. The "models" in the mimicry system are the bad sub-types of the previous section, and the "mimics" are the good sub-types. Estabrook and Jespersen were interested in the conditions for the model-mimic system to be ignored or exploited by the predator. Arnold(1978) considered the same model from the more general perspective of "optimal modifiable behaviour."

Like the model of section 8.2, the Estabrook-Jespersen-Arnold model has three components. The forager simultaneously encounters an ambiguous prey type X , and an alternative prey type of constant value. Prey type X is composed of two sub-types, good and bad. The principal difference between this model and that of section 8.2 is that it considers all possible changes from one encounter to the next: good to good, good to bad, bad to bad, bad to good. Specifically, let g_i be the event that encounter i is with a good, similarly b_i signifies that encounter i is with a bad. The general Estabrook-Jespersen-Arnold model allows for a arbitrary Markov transformation matrix of the form

$$\begin{array}{cc} & \begin{array}{cc} g_{i+1} & b_{i+1} \end{array} \\ \begin{array}{c} g_i \\ b_i \end{array} & \begin{bmatrix} q_g & 1-q_g \\ 1-q_b & q_b \end{bmatrix} \end{array}$$

where

$$\begin{aligned} q_g &= P(g_{i+1} \mid g_i) \\ 1 - q_g &= P(b_{i+1} \mid g_i) \\ q_b &= P(b_{i+1} \mid b_i) \\ 1 - q_b &= P(g_{i+1} \mid b_i) \end{aligned}$$

For simplicity I restrict my attention to the case where $q_g = q_b = q$. q may still be called the stationary probability, as in 8.2, but in this case it applies to runs of good luck as well as to runs of bad luck. Secondly, at its most general the Estabrook-Jespersion-Arnold model allows for $0 \leq q \leq 1$. In the case where $q < 1/2$, the forager gains "negative" information about prey type X , that is, the sub-type is more likely to change than to stay the same. I do not think there are many plausible natural situations where this kind of good-bad alternation will occur. Consequentially I restrict my attention to the positive information case, $q > 1/2$.

Since this system retains the Markov property, and good and bad sub-types are immediately recognized once eaten, we still seek an optimal sampling period, N^* . Let $p_g(N)$ be the probability of eating a good given that a sampling period of N is adopted. Similarly, let $p_b(N)$ be the probability of eating a bad. The expected food gain per encounter is given by:

$$v_b p_b(N) + v_g p_g(N) + v_a \{1 - p_b(N) - p_g(N)\} \quad (8.8a)$$

which becomes

$$(v_g - v_a) \{p_g(N) - b p_b(N)\} + v_a \quad (8.8b)$$

As long as $v_g - v_a > 0$, the expected gain per encounter may be maximized by maximizing

$$p_g(N) - b p_b(N) \quad (8.9)$$

in the choice of N .

Estabrook and Jespersen (1974) have elegantly derived expressions for the equilibrium probabilities $p_g(N)$ and $p_b(N)$, (they are "equilibrium" in the sense that after many encounters a forager adopting $N = N_1$ will have eaten goods on proportion $p_g(N_1)$ of

encounters and bads on proportion $p_b(N_1)$ of encounters). Define $\kappa = 2q-1 = 1-2p$, where p is the transition probability $1-q$. By substituting Estabrook and Jespersen's expressions (3) and (4), which give $p_g(N)$ and $p_b(N)$ respectively, into 8.9. I find that the problem of maximizing expected gain per encounter is equivalent to maximizing

$$f_N = \frac{(1 - \kappa^{N+1}) - b(1 - \kappa)}{(1 - \kappa)(N + 1) + (1 - \kappa^{N+1})} \quad (8.10)$$

The analysis proceeds almost identically to that of section 8.2. I want to maximize f_N in N , so I seek conditions for $f_N - f_{N-1} > 0$, ($N \geq 1$). Solving for a condition on the error ratio b , I find that f_N is increasing with N if

$$b > \frac{s_N - N\kappa^N}{1 + \kappa^N} = b^*(N, q), \quad (8.11)$$

where s_N is the sum of a geometric series with first term equal to one and common ratio κ [i.e. $s_N = (1 - \kappa^{N+1})/(1 - \kappa)$]. It is straightforward to show that $b^*(N, q)$ always increases with N , since

$$b^*(N, q) - b^*(N-1, q) > 0 \quad (8.12a)$$

is equivalent to the condition

$$(1 - \kappa)(s_N + N) > 0 \quad (8.12b)$$

which will always be true, because $0 < \kappa < 1$, and $N \geq 1$. Since $b^*(N, q)$ increases monotonically in N it is easily to find the conditions for which a non-trivial, i.e. $1 \leq N^* < \infty$, sampling period should be chosen:

$$\frac{1 - \kappa}{1 + \kappa} < b < \frac{1}{1 - \kappa} \quad (8.13)$$

More specifically, the prey type X should always be eaten, $N^* = 0$, if

$$b < \frac{1 - \kappa}{1 + \kappa} = \frac{1 - q}{q} = \frac{1}{E(L)},$$

and, prey type X should always be ignored ($N^* = \infty$) if

$$b > \frac{1}{1 - \kappa} = \frac{1}{2(1 - q)} = \frac{1}{2}[E(L) - 1]$$

The special case considered here ($q_g = q_b$, and $q > 1/2$) allows me to

specify stricter conditions than those found by Estabrook and Jespersen (1974), Bobisud and Potratz (1976) and Arnold (1978).

Within the non-trivial solution region given by (8.13), we wish to find the optimal N . Reconsider expression (8.11) and $b^*(N,q)$. In figure 8.3 I plot isoclines of $b^*(N,q)$ for a number of values of N , and at each value of N over a wide range of q 's. These curves have the following interpretation. At a fixed q , if b is between $b^*(N,q)$ and $b^*(N+1,q)$ then N is the optimal sampling period. Therefore the region between the curves $b^*(2,q)$ and $b^*(3,q)$ delineates the conditions of b and q for which two is the optimal sampling period, $N^*=2$. Figure 8.3 represents a graphical solution to finding the optimal sampling period.

The figure shows several interesting features. Firstly, notice that the region where non-trivial sampling strategies are optimal is actually quite a small part of the parameter space. For q between $1/2$ and $3/4$ non-trivial N^* are very improbable. Sampling behaviour is most likely when q is near one. (q near one is the condition the run-of-bad-luck model of section 8.2 approximates.) At a fixed stationary probability (a slice parallel to the b axis of figure 8.3) the optimal N is shown to be an increasing function of b , the error ratio. Figure 8.4 shows such a slice of figure 8.3 at $q = 19/20$; figure 8.4 gives the optimal sampling period as a function of b for this condition. As shown in section 8.2, optimal sampling period is an increasing function of the error ratio, b . In this case, the optimal sampling period can be seen to increase asymptotically as b approaches $1/(1-\kappa) = 10$.

A third, rather puzzling, result is evident from figure 8.2. Consider the effect of increasing the stationary probability q at a fixed b , where b is slightly greater than 1. At q near one-half,

Figure 8.3 A General Graphical Solution of Optimal Sampling Period.

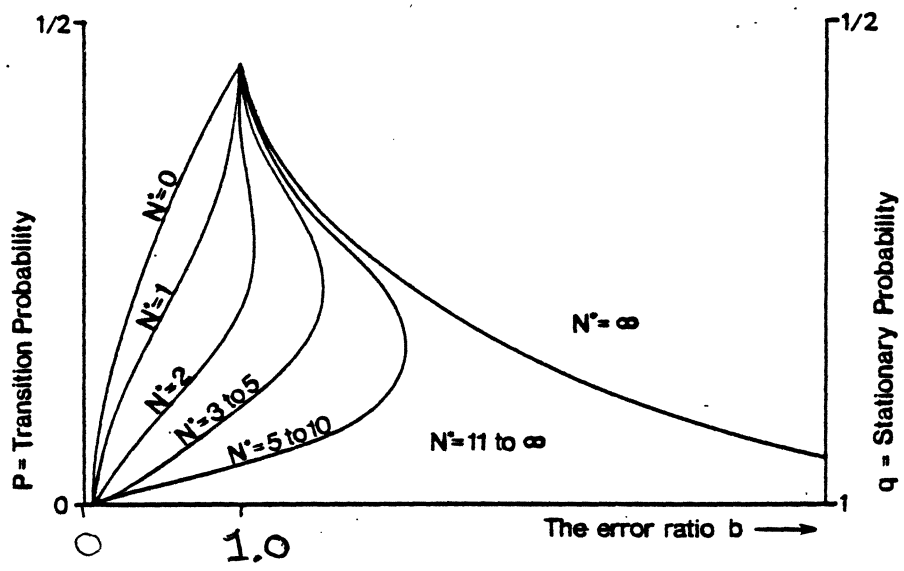


Figure 8.3 The figure shows the regions of q (stationary probability) and b (error ratio) where various optimal sampling periods obtain. Note the restrictive nature of the conditions for N^* to be different from 0 or infinity.

Figure 8.4 Optimal Sampling Period as a function of Error Ratio.

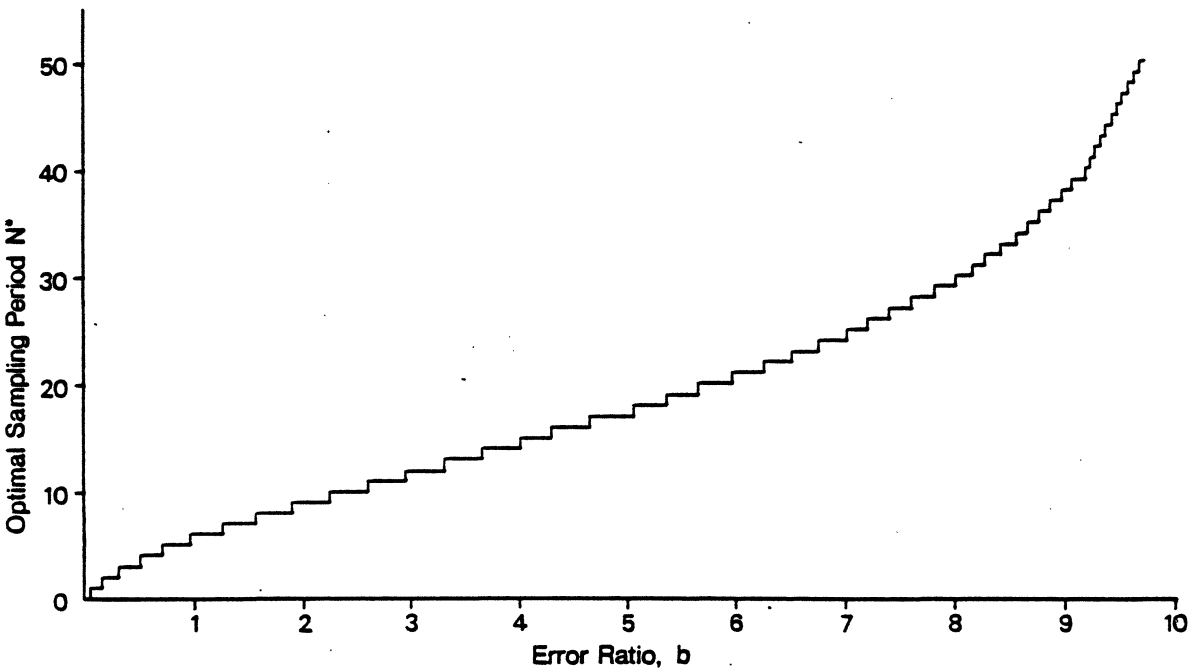


Figure 8.4 The figure shows optimal sampling period as a function of b the error ratio, initial condition $q=19/20$, calculated from the more general model of section 8.3

$N^* = \infty$, and as q increases N^* decreases, but as q increases further N^* begins to increase. I do not have a convincing explanation of this predicted phenomenon, but I do have one possible explanation. In other words, this effect is: when $b > 1$ the smallest N^* 's are at values intermediate between $q=1/2$ and $q=1$. Small N^* 's reflect careful tracking of the environment. N^* should be large if q is near one because the forager is confident of a fairly long run of bad luck. When q is near $1/2$, the forager knows that runs of bad luck are likely to be short, but so are runs of good luck. So that the value of finding a "good" is lower (because the run of good luck it foretells is shorter) when q is near $1/2$ than when it has an intermediate value. This interpretation is similar to McNamara's (1982) concept of the "potential value of information."

8.4 Discussion and Summary

Since this chapter has principally been a re-analysis of a previously presented model I outline here precisely what is new, or different, about my approach. I have tried to expand the Estabrook-Jespersion-Arnold model in two directions. Firstly, I have tried to make its results more intuitively appealing and more analytical, and secondly, I have modified its form to make it more readily applicable to the experimental situation described in the next chapter.

My interpretation of b as the ratio of the cost of sampling errors to the cost of over-run errors is new, and it has two advantages. Firstly, it extends the model to the more general foraging case, as opposed to the model-mimic system discussed by Estabrook and Jespersen. Secondly, it brings the value of alternative prey items explicitly into the formulation; while this is not mathematically

important, it is important for understanding the biological situation envisioned, as well as for testing the model. The names given to b by Estabrook and Jespersen (1974) and Arnold (1974) are potentially confusing. "Relative noxiousness" (Estabrook and Jespersen) and "relative badness" (Arnold) could easily be misinterpreted to mean v_b/v_g . Considering the alternatives only implicitly can lead to confusion. For example, Arnold considers the $N^* = \infty$ case to be different from the "always ignore the X-prey type" case. This is an artefact of maximizing only the gain due to encounters with the X-prey type (i.e. tacitly ignoring alternatives). In fact $N^* = \infty$ can be optimal ^{only} if the expected gain due to encounters with the X-prey type is less than or equal to zero, the same case in which [#]the X-prey type should always be ignored.

By considering only the positive information case, I am able to obtain stricter conditions for non-trivial sampling strategies [my condition for the "always ignore" strategy to be optimal-- $b > 1/(1-\kappa)$ --is one half the size of Estabrook and Jespersen's-- $b > 2/(1-\kappa)$] and a more analytical solution than either Estabrook and Jespersen's boundary conditions, or Arnold's computational results. Figure 8.3 represents a significant advance towards an analytical solution to even this simple environmental tracking problem.

Summary of Results. The simple model for optimal treatment of a run of bad luck in section 8.2 is intended to illustrate the importance of the error ratio, b , in problems of environmental tracking. b re-appears in the more general model of section 8.3, and I expect that b , or a similar term, will be of general importance in models concerned with optimal sampling frequency. The trade-off between sensitivity to change and the cost of sampling is of general importance. In section

8.3 I show that the optimal sampling period, N^* , is an infinitely increasing function of b , figure 8.4. I test this result experimentally in the next chapter.

A second important result is the narrowness of the conditions for which sampling is a sensible strategy, a fact also commented on by Estabrook and Jespersen (1974) and Arnold (1978). The conditions for sampling to be optimal are quite special. Sampling is only useful, broadly speaking, when the stationary probability (q) is near one. Moreover, consider the effects of changing the assumptions of the model. There are two dubious assumptions: that recognition of sub-types takes only one sample, and that the system changes according to a Markov rule. Violating either of these assumptions will have the effect of making sampling more expensive, and consequently of limiting the region of plausible sampling strategies even further. I think it is fair to conclude that in nature sampling a changing environment will only be valuable when "runs" are very long (q approaches 1). The idea that sampling, in general, is valuable over very limited conditions is discussed further in chapter ten.

Finally, the chief importance of this re-analysis is its application to the experimental results in the next chapter, in which I test the hypothesis that sampling period increases with the error ratio b , as shown in figure 8.4

CHAPTER NINE

Tracking a Changing Environment: a Test

9.1 Introduction

In this chapter I present an experimental test of the theoretical results of chapter eight. Specifically, I vary the ratio of the cost of sampling errors to the cost of over-run errors (b), and I test whether changes in b coincide with the predicted changes in the sampling period (N^*). Statistically, I ask two questions: Is the change in sampling between treatments as predicted? and, is the amount of sampling in each treatment as predicted?

I pointed out in chapter eight that the simplifying and unnatural assumptions made by this model of environmental tracking will tend to make sampling more likely. Therefore it is reasonable to expect that the model will predict the trend in sampling more accurately than the amount of sampling.

Although chapter eight's model of environmental tracking was developed in terms of encounters with a prey item, the experiment presents the forager with a series of dichotomous choices between patches. One patch changes according to the symmetric first-order Markov process described in chapter eight (the changing patch), and the other provides a constant probability of reward (the alternative or constant patch). Both patches provide either a single reward (fly pupa) or no reward at a given encounter. Recall (as described in chapter eight) that the definition of b is

$$b = \frac{v_a - v_b}{v_g - v_a}$$

where v_a is the value of the alternative(s) (the constant patch), v_b is the value of the bad sub-types in the changing patch and v_g

is the value of good sub-types in the changing patch. In the experimental case $v_a = r_a$ = the probability of reward in the constant patch, and $v_g = 1$ and $v_b = 0$. Therefore,

$$b = \frac{r_a}{1 - r_a}$$

In the experiment I change b by changing r_a , the probability of reward in the constant patch.

9.2 Materials and Methods

Subjects and their care. The experimental subjects were three great tits (Parus major), two females (birds one and two) and one male (Bird three). The birds were kept individually in large outdoor aviaries. Each outdoor aviary had a small door which communicated with the indoor experimental aviary; this door could be opened and closed by the experimenter unobserved by the birds. The aviaries used were the same as those described by Kacelnik (1979). The birds were fed a diet of commercial food (Aleckwa Mischung I), with added boiled egg and grated carrot, normally supplemented with peanuts and sunflower seeds. During the experiment the food was not supplemented with nuts. The experimental sessions were always conducted between 0830 hrs. and 1230 hrs. The birds were given their standard food 15 to 30 min. after finishing their daily experiments. Food was removed from the outdoor aviaries at 1700 hrs. or dusk whichever came first, so that the birds would be hungry for the experimental sessions the following morning. Experiments were normally conducted six days per week; so that the birds could have a rest day without food deprivation to avoid cumulative food stress. Nuts were provided on the rest day.

Hardware: the aviary equipment

A diagram of the experimental aviary is shown in figure 9.1. The elements of the aviary are: (1) two feeding discs each fitted with a

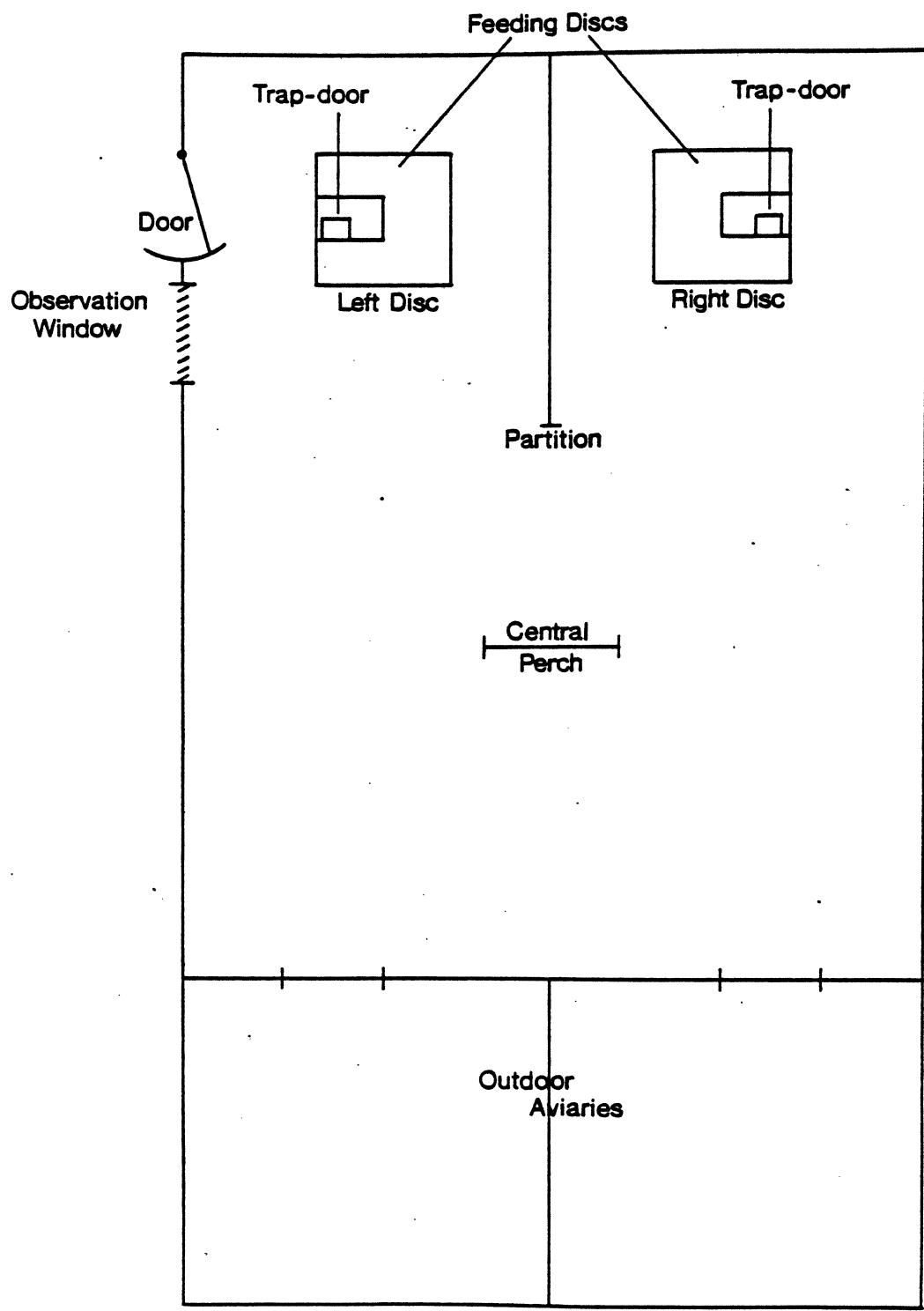
photocell operated trapdoor, (2) a central electronic perch and (3) a wire mesh partition separating the two discs.

The discs. The birds were trained to feed from two remotely controlled feeding discs as described by Kacelnik (1979). Each disc contained 5 rows of 72 holes. The experimental rewards were fly pupae placed in these holes. The discs operated by sequentially presenting the reward holes. I refer to this presentation as a "disc advance," the discs were advanced only after a bird hopped on the central perch. The discs were always advanced simultaneously. I distinguish between the discs by calling them right and left as shown in figure 9.1

The trapdoors. Each disc was fitted with an opaque trapdoor placed over the food presentation slot of the disc. The trapdoor had two functions. Firstly, it prevented the bird from seeing whether or not a reward was present without sampling. Secondly, it recorded when sampling (a "look") occurred. The trapdoors were fitted with a photocell. When the bird broke the light-beam with its head (this only happened when its head was directly over the reward slot) a solenoid was activated which opened and held open the door for an adjustable period of time. Moreover, it was possible to prevent the trapdoors from opening electronically by switching them into "inhibit mode."

The central perch. Between the two feeding discs I placed an electronically monitored perch, (see figure 9.1 for placement). The perch was operated by a microswitch, and allowed me to tell remotely whether or not the bird was on the perch. This perch was positioned so that it was equidistant from both discs. This positioning insured that each choice was independent of previous choices; because going to either disc had the same energetic costs.

Figure 9.1 Plan of the Experimental Aviary.



(not to scale)

Experimental control and data collection

The experimental equipment described above was connected to devices which allowed me to remotely control the experiment and record data. The experiment was controlled by commercially available logical equipment, collectively called "the logic". Data were recorded by a Digital PDP-8e computer and punched on to paper tape for later analysis.

Programming the logic

The logic was programmed to do the following. (1) The discs were advanced by a "hop" on the central perch. The first hop registered after a "look" was the only hop to cause a disc advanced, i.e. the first hop inhibited the perch. (2) A look, registered by the photocell of one of the trapdoor^s disinhibited further looks (3) A hop following a look disinhibited the trapdoors. The birds quickly learned this scheme of "hop-look" alternation.

The following were recorded by the computer. (1) Effective hops (those which advanced the discs), (2) Effective looks (those which opened a trapdoor) and which disc (left or right) was visited. Three events were recorded, perch hops, ^{right} disc looks, and ^{left} disc looks. The computer also recorded the time of each event.

Experimental training and treatments.

The birds were admitted to the experimental aviary twice a day, i.e. there were normally two "sessions" per bird per day. Each session consisted of 72 hop-look cycles; so that an entire row of each disc was presented in each session.

Two types of reward schedule were used in each treatment, a changing schedule and a constant schedule. The changing schedules were ten schedules of 72 presentations each which were the same for all birds in all treatments. Only one changing schedule was presented per session. The starting point in the cycle of ten was randomly chosen

for each bird in each treatment. The changing schedules were generated, using computer produced random numbers, by applying the first-order Markov rule described in section 8.3 with stationary probability, q , set equal to 0.9863. This q was chosen so that the expected run length was equal to 72 disc advances, $[q/(1-q)=72]$. This meant that on average there was one transition (good to bad, or bad to good) per session. This rather peculiar number was chosen to enhance the difference between sampling strategies and trivial "stick-on-one-side" strategies (this difference gets bigger as the run length gets longer). The "good" sub-types were holes with a fly pupa in them and the "bad" sub-types were empty holes.

The constant schedules were different for each treatment. The constant disc provided fly pupae at an unchanging probability per hole within each treatment (VR schedule). Three probabilities of reward (r_a 's) were used, 0.2, 0.5, and 0.8. These treatments correspond to error ratios of $b=0.25$, $b=1.0$, and $b=4$ respectively. There were 6 schedules of 72 for each treatment, generated by computer. The starting point in the cycles of six was determined at random for each bird in each treatment. The order of treatments was randomized for each bird.

The treatment cycle The model of chapter eight assumes that the forager knows the characteristics of each prey type. Therefore, I included as part of each treatment a period of "treatment training" to teach the birds the characteristics of its environment. This training used the inhibit facility of the trapdoors to force the birds to alternate every ten hops between the right and left discs. These sessions are called "forced choice" sessions. Each error ratio treatment followed the treatment cycle listed below and in table 9.1

TABLE 9.1 THE TREATMENT CYCLE

DAY	SESSION	TYPE OF TREATMENT	CHOICE REGIME
1	1	NEUTRALIZATION	FORCED
2	2		
	3		
	4		
3	5	PREFERENCE TEST	FREE
	6		
4	7	TREATMENT TRAINING	FORCED
	8		
5	9		
	10		
6	11		
	12		
7	13		
	14		
8	15		
	16		
9	17	EQUILIBRATION	FREE
	18		
10	19	OBSERVATION	FREE
	20		
11	21		
	22		
12	23		
	24		
13	25		
	26		
14	27		
	28		

Days 1 and 2, Sessions 1 to 4, Neutralization The bird was given a low constant VR schedule (10% to 25% reward) on both discs and forced choices.

Day 3, Sessions 5 and 6, Preference test The bird was given the same types of schedules as above, but free choices. The birds always showed a numerical preference for either the left or right disc. The results of this test determined which disc was the changing disc in the remainder of the treatment. These assignments were made at random with respect to preference.

Days 4 to 8, Sessions 7 to 16, Treatment Training The bird was given the experimental schedules and forced choices for ten sessions.

Day 9, Session 17 and 18, Equilibration As above, but free choice. The bird was allowed to get used to having free choices again. Data were recorded, but were excluded from the analysis.

Days 10 to 14, Sessions 19 to 28, Observation. The bird was given experimental schedules and free choices. Data were recorded, and used in the analysis.

Data In the analysis of the results I have used the three levels of analysis of operant behaviour proposed by Staddon (1980). Firstly, I have asked questions of Molar description; is the overall sampling pattern as expected? Secondly, I have investigated performance; do the birds achieve feeding rates at or near the optima? Thirdly, I consider a molecular analysis of the behaviour; how well does the observed behaviour actually fit the simple-minded, fixed sampling period rule proposed?

Five sets of data were recorded for each bird in each treatment, although bird three died after its first treatment (80% reward). These were: (1) sampling periods (recall that period approx. $1/\text{sampling}$

frequency, not the duration of the sampling event), where a sampling period is the number of encounters (Looks) spent away from the changing disc (i.e. feeding on the constant disc) after the bird observes an empty hole on the changing disc. This variable can only take the values, 0,1,2,...etc. (2) The proportion of encounters spent feeding on the alternative (constant) disc per session. This is a convenient variable for statistical analysis and the theory makes direct predictions about its magnitude. The last three remaining variables recorded are chiefly of interest in the molecular analysis of the predicted sampling rule: (3) the overall rate of pupae gained per encounter; (4) the proportion of win-shift errors, i.e. errors of leaving the changing side after being rewarded (which the theory does not allow); (5) the proportion of lose-stay errors, i.e. errors of staying on the changing side after not being rewarded. The predicted values for these observations are summarized in table 9.2

9.3 Results: molar description

Sampling Period. Figure 9.2 presents the predicted relationship between optimal sampling period and the error ratio b , and the observed means ($\pm$ s.e.) of sampling period observed (data for all birds). The figure shows that the means fit the predicted patterns well. Specifically, the predicted negative acceleration of sampling period with b fits the data well. These data are summarized in table 9.3, where they are divided by treatments and individuals. Birds one and two (those with complete data sets) differ most markedly in $b=1 / r_a=0.5$ treatment.

Figure 9.3 shows histograms of the frequency distributions of sampling period for all birds grouped by treatment. Sampling periods of zero were the mode in all treatments. These data show a strongly

TABLE 9.2 EXPERIMENTAL PREDICTIONS

TREATMENTS		PREDICTIONS				
r a	b	Optimal Sampling Period	Pro. on Constant Side	Optimal Rate R*	Pro. Win- Shift	Pro. Lose- Stay
0.2	0.25	5	0.431	0.604	0	0
0.5	1.0	11	0.492	0.752	0	0
0.8	4.0	25	0.572	0.866	0	0

Figure 9.2 Predictions and Observations of Sampling Period

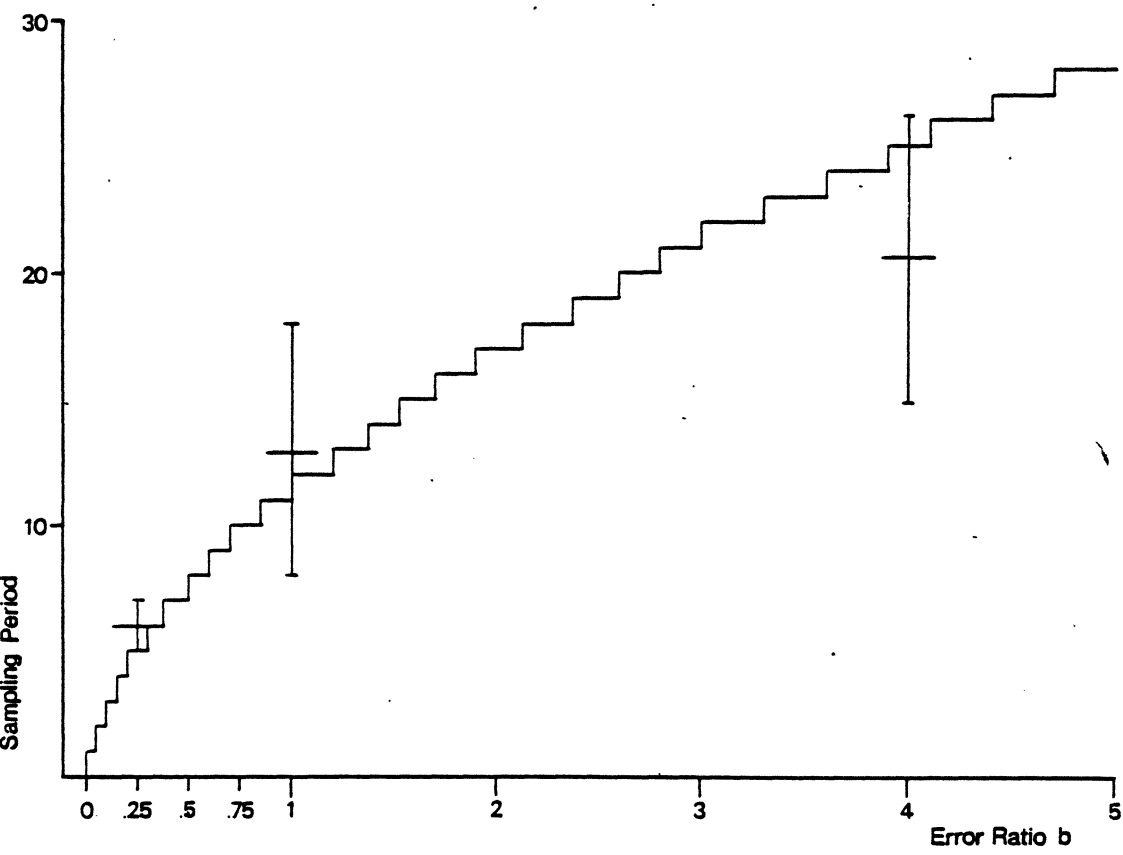


Figure 9.2 The figure shows the predicted curve for sampling period as a function of b, the Error Ratio, and the observed mean standard error bars.

TABLE 9.3 SUMMARY OF SAMPLING PERIOD DATA

TREATMENT		BIRD(S)	$\bar{N} \pm \text{s.e.}$	SAMPLE* SIZE
r a	b			
0.2	0.25	PREDICTED	5	
		ALL	6.0 ± 0.98	156
		ONE	5.1 ± 0.82	92
		TWO	7.4 ± 2.0	64
0.5	1.0	PREDICTED	11	
		ALL	11.9 ± 5.1	48
		ONE	48.0 ± 37.8	6
		TWO	6.6 ± 1.8	42
0.8	4.0	PREDICTED	25	
		ALL	20.6 ± 5.4	58
		ONE	17.8 ± 6.3	25
		TWO	10.7 ± 3.8	27
		THREE	67.5 ± 39.8	6

* Sample size is the number of sampling periods observed

Figure 9.3 Histograms of Sampling Period divided by treatments.

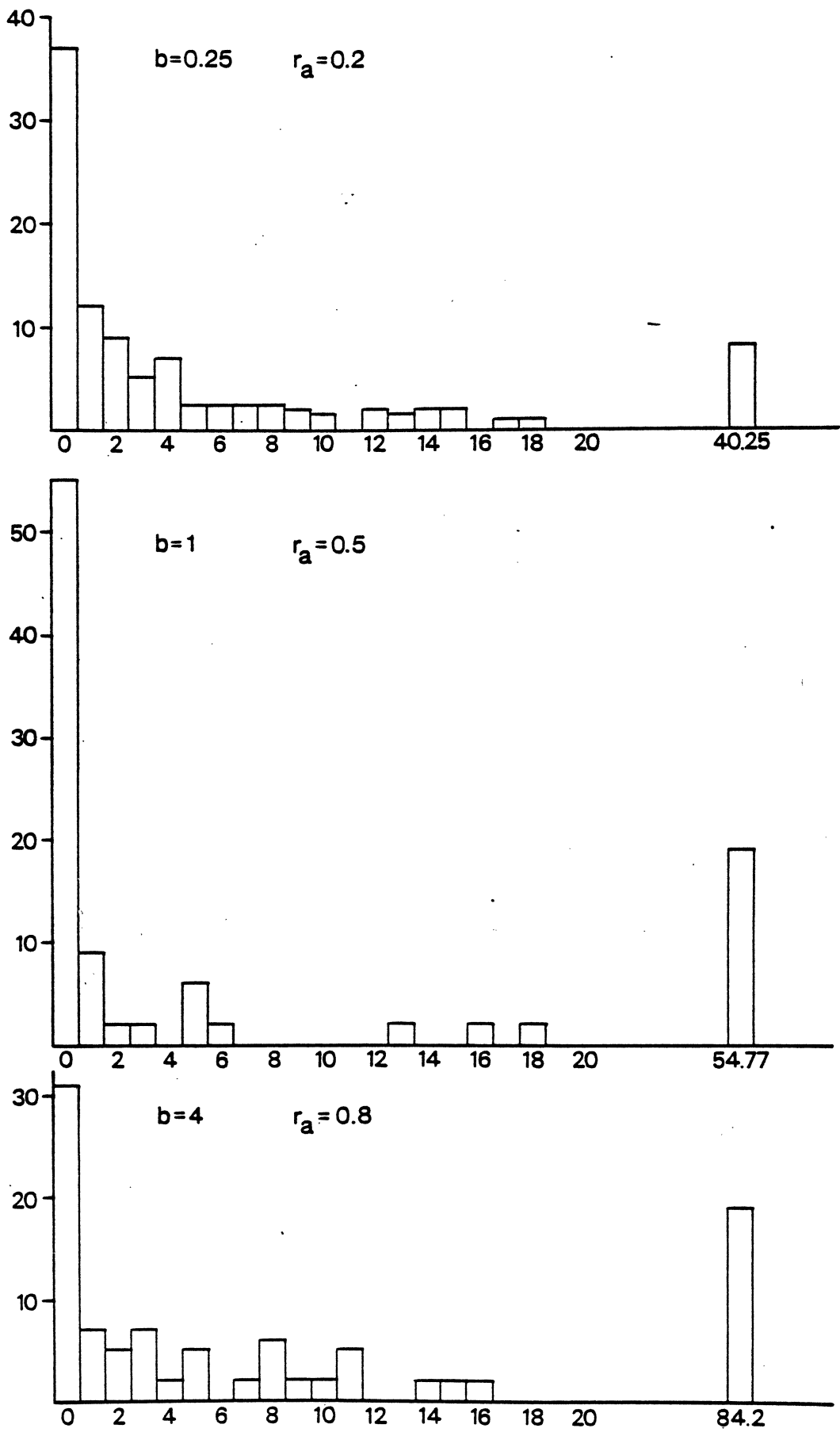


Figure 9.3 Histograms of sampling period (A) $b=0.25$ treatment; (B) $b=1$ treatment and (C) $b=4$ treatment. The last bar in each histogram lumps the observations greater than twenty, and shows the mean value of these observations.

non-normal distribution; consequently, my statistical analyses of sampling period are non-parametric.

Specifically, I test the idea that sampling period increases with error ratio. The results of these calculations are presented in table 9.4. All the correlations are positive, although the correlation for bird two alone is insignificantly so. The data support the hypothesis that sampling period increases with the ratio of the cost of sampling mistakes to the cost of over-run mistakes, b . This evidence is only qualitative, although figure 9.2 suggests a more quantitative fit.

Proportion of "time" spent on the constant side The theory of chapter eight also predicts the proportion of "looks" which the birds should spend at the constant disc (table 9.2). Specifically, the theory predicts that more time should be spent in the constant side (which corresponds to less sampling) as the error ratio increases. This is a convenient variable for statistical analysis because techniques are available for transforming proportions to meet the assumptions of parametric statistics.

To test for differences in the means for different treatments and different birds I performed a two-way analysis of variance on the arcsin transformed data, using the BMD computer package for two-way analyses of variance with unequal cell sizes (Dixon 1979). The results show a significant interaction between individual and error ratio treatments ($p < 0.05$), as well as significant effects of individual and treatment. Table 9.5 shows the cell means for this analysis. The table shows that the proportions are very similar for birds one and two in the $b=0.25$ and $b=4$ treatments, but very different for the $b=1$ treatment. I performed a second analysis of variance without the $b=1$ treatment and I found that the results were in accord

TABLE 9.4 SPEARMANS RANK CORRELATION COEFFICIENTS (r_s):
TREATMENT VS SAMPLING PERIOD

DATA	r_s	t	Sample ¹ Size
ALL BIRDS	0.1123	1.82*	262
BIRD ONE	0.4104	4.97*	123
BIRD TWO	0.0554	0.68(ns)	133

¹ Sample size is the number of sampling periods observed. * indicates that r_s is significantly different from zero, $p < .05$.

TABLE 9.5 CELL MEANS FOR ANALYSIS OF VARIANCE:
ARCSIN TRANSFORMED PROPORTION OF TIME ON CONSTANT SIDE VS. INDIVIDUALS

TREATMENT (b)			
	b = 0.25	b = 1.0	b = 4.0
BIRD ONE:	0.60065 (14)	1.4276 (11)	1.1025 (10)
BIRD TWO:	0.76765 (12)	0.73364 (12)	1.09224 (8)

Sample sizes are show in parenthesis. The sample size is the proportion of time on constant per session.

TABLE 9.6 REGRESSION STATISTICS:
REGRESSION OF ARCSIN TRANSFORMED PROPORTION OF TIME ON
CONSTANT VS. REWARD ON CONSTANT SIDE

DATA	EQUATION	SAMPLE SIZE	CORRELATION COEFFICIENT	DIFF FROM ZERO?	F -LINEAR	Df
ALL BIRDS	0.896X + 0.535	77	0.497	p < 0.05	1.46	1,74
BIRD ONE	0.939X + 0.567	35	0.489	p < 0.05	22.2*	1,32
BIRD TWO	0.491X + 0.609	32	0.296	p 0.1	1.9	1,29

Sample size is the no of sessions in which the proportion was observed; one proportion per session. * indicated a significant p <0.05 departure from linearity. The fifth column reports the results of a test of whether the correlation coefficient is different from zero.

with the theoretical predictions; the results showed insignificant interaction and individual effects, but significant effects of b , ($p < 0.05$). I consider the anomalous treatment for bird one at $b=1$ in the discussion. I conclude that there are real differences between the treatments.

For the three treatments tested the Arcsine of the proportion of time on the constant side is predicted to increase approximately linearly with reward probability on the constant side, r_a . To test this I performed linear regressions. The results of these calculations are presented in table 9.6. The table shows regression statistics for all birds (one, two, and three), bird one individually, and bird two individually. All the correlation coefficients are significantly different from zero (although bird two's is less so). I tested the regressions for linearity using an analysis of variance technique described by Dixon and Massey (1969; the F-ratios for this analysis are given in column six of table 9.6). The data for bird one alone was found to depart significantly from linearity ($p < 0.05$). These results support not only the existence of the predicted trend, but they support the form in the increase as well. Figure 9.4 shows the predicted and observed lines (data for all birds). It is immediately evident from the figure that in all treatments the birds spent much more time on the constant side than predicted (data summarized in table 9.7), and that this is increasingly so as the probability of reward on the constant side (r_a) increases. Although the predicted trend is found, the birds sample less than predicted in every case.

Results--Performance: how well do the birds do?

In table 9.8 I compare the rate of food gain (measured in pupae per "look") with the optimal rate, the highest rate obtainable by

Figure 9.4 Linear Regression of transformed proportion of time on constant side vs. reward on constant side.

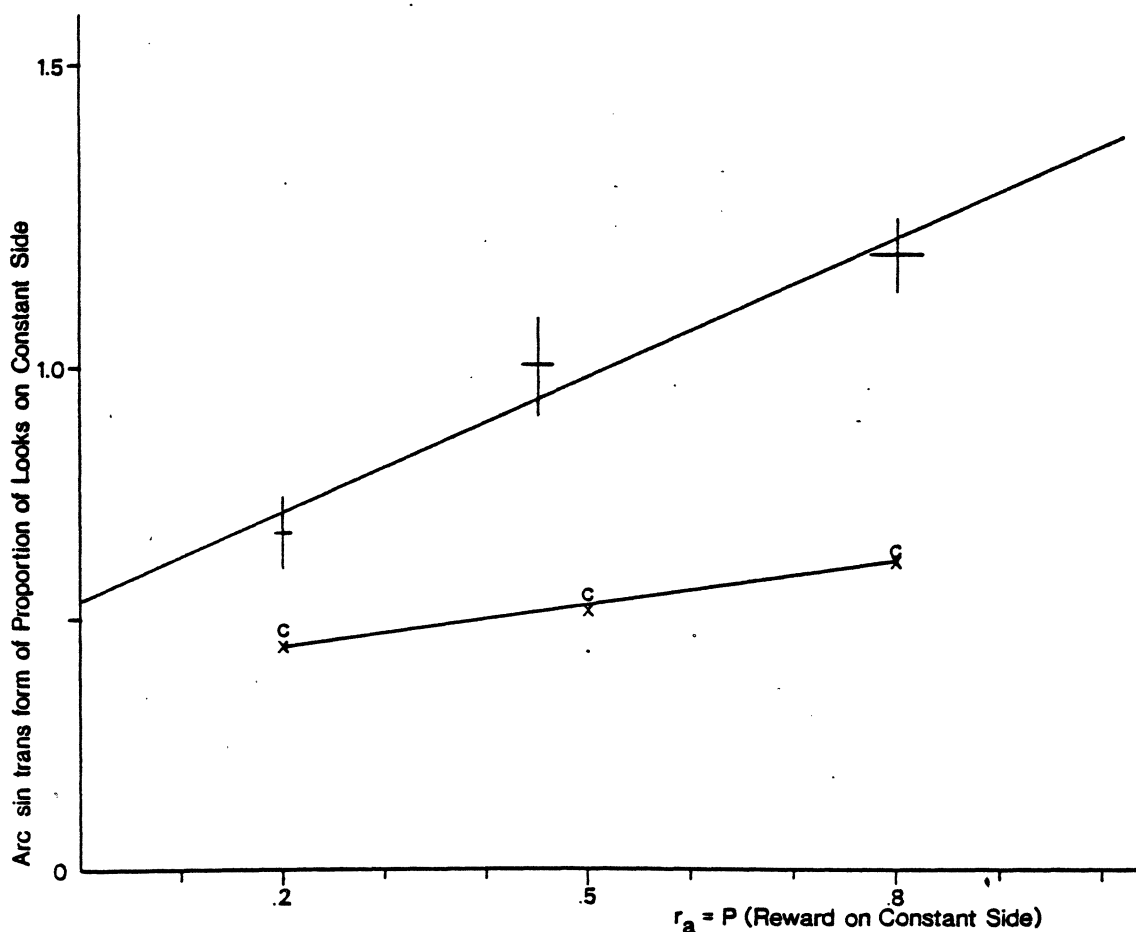


Figure 9.4 The figure shows the probability of reward on the constant side on the abscissa and the Arcsine transformed proportion of time on the constant side. The upper line is that estimated from all data shown with means + standard error bars. The lower line marked with crosses is the predicted line. The birds seem to spend more time on the constant side than predicted. The C's represent predictions corrected for sampling clump size.

TABLE 9.7 DATA SUMMARY: PROPORTION OF TIME ON THE CONSTANT SIDE.

UNTRANSFORMED DATA

TREATMENT		PREDICTIONS		DATA			
r	b	Random	Optimal	All Birds	Bird One	Bird Two	Bird Three
a							
0.2	0.25	0.5	0.431	0.584	0.526	0.652	na
0.5	1.0	0.5	0.492	0.801	0.987	0.636	na
0.8	4.0	0.5	0.572	0.884	0.849	0.826	0.971

choosing the disc with the higher average reward rate for that treatment (such a strategy would require some sampling) and the rate due to random choice. The achieved rate always lies between random choice and the optimal rate, but it is only greater than the maximum one disc strategy labelled $[\text{Max}(\text{Ch.}, \text{Con.})]$ for the $b=1, r_a=0.5$ treatment. I conclude that the birds are sampling. They are using information to do better than random, but they are not doing so by simply remaining on the "best" single disc.

Results--Molecular Analysis: Do the birds use a sampling period rule?

In table 9.9 I show data on the proportion of win-shift (leaving the changing side after a reward) and lose-stay (staying at the changing side after a failure) mistakes made by the birds. If the birds had actually adopted a fixed sampling period, then they should have made none of these mistakes. In fact, they do seem to make very few win-shift mistakes, but they do make a surprising number of win-shift mistakes in the $b=4, r_a=0.8$ treatment. They are much worse at lose-stay mistakes. This hints at the weakness of the assumption of immediate recognition of a run of bad luck. Finally, if the birds had adopted a sampling period rule then they should never have taken two empty holes in a row from the variable disc. Sampling should not occur in "clumps". In table 9.10 I summarize data on clump size (number of "looks" in a row). For all birds the mean clump size is 1.59 ± 0.082 [s.e.(N=170)]. Again this suggests that immediate recognition may be an unnatural assumption. If this is so, it would affect my estimate of the error ratio (b) as a function of reward probability r_a (which is central to this experimental design). For example, imagine that the birds act under the assumption that on average 1.56 samples are required to be sure that a run of bad luck is in progress. This would

TABLE 9.8 SUMMARY OF RATE DATA: PERFORMANCE.

TREATMENT		PREDICTIONS			DATA			
r a	b	Max(Ch.,Con)	Random	Optimal	All Birds	Bird One	Bird Two	Bird Three
0.2	0.25	0.5	0.35	0.604	0.45	0.48	0.41	na
0.5	1.0	0.5	0.5	0.752	0.55	0.48	0.62	na
0.8	4.0	0.8	0.65	0.866	0.79	0.80	0.78	0.81

TABLE 9.9a SUMMARY OF MISTAKE DATA: WIN-SHIFT MISTAKES

TREATMENT		PREDICTION	DATA			
r	b		All Birds	Bird One	Bird Two	Bird Three
a						
0.2	0.25	0.0	0.065 (615)	0.065 (372)	0.062 (243)	na
0.5	1.0	0.0	0.058 (242)	1.0 (1)	0.054 (241)	na
0.8	4.0	0.0	0.164 (189)	0.220 (82)	0.130 (92)	0.067 (15)

Numbers in parenthesis indicate the sample size on which the proportions are based; the number of times a win was observed.

TABLE 9.9b SUMMARY OF MISTAKE DATA: LOSE-STAY MISTAKES

TREATMENT		PREDICTION	DATA			
r	b		All Birds	Bird One	Bird Two	Bird Three
a						
0.2	0.25	0.0	0.353 (153)	0.247 (89)	0.500 (64)	na
0.5	1.0	0.0	0.411 (56)	0.25 (8)	0.438 (48)	na
0.8	4.0	0.0	0.274 (62)	0.259 (27)	0.130 (29)	0.000 (6)

Numbers in parenthesis indicate the sample size on which the proportions are based; the number of times a loss was observed.

TABLE 9.10 DATA SUMMARY: "CLUMP SIZE"

MEANS $\pm$ STANDARD ERRORS
TREATMENT

	$r = 0.2$ a	$r = 0.5$ a	$r = 0.8$ a
ALL BIRDS	1.57 ± 0.103 (100)	1.93 ± 0.270 (28)	1.43 ± 0.124 (42)
BIRD ONE	1.39 ± 0.110 (70)	1.40 ± 0.330 (5)	1.35 ± 0.202 (20)
BIRD TWO	2.00 ± 0.215 (30)	2.00 ± 0.260 (23)	1.69 ± 0.234 (16)
BIRD THREE	na	na	1.00 ± 0.00 (6)

Numbers in parenthesis are sample sizes; the number of clumps observed.

have the effect of increasing b by a factor of 1.56. This change would predict less sampling, but it is not enough to make much difference in figure 9.3 where I have plotted these "corrected" predictions as "C's". This suggests that birds still spend far too much time at the constant disc than the model predicts.

9.4 Discussion and summary

I deduce three main points from the experimental results. Firstly, the degree and frequency of sampling changes with the ratio of the cost of sampling errors to the cost of over-run errors (b) in the predicted way, (figures 9.2 and 9.4). Specifically, the frequency of sampling decreases as b , the error ratio, increases. Secondly, the absolute amount of sampling is much lower than predicted by the model in chapter eight. I deduce this by comparing the linear regression of figure 9.4 with the predicted line and also from the large number of zero sampling periods observed (figure 9.3). In general, I conclude that the birds did not track their environment as carefully as predicted, and that they could have done better (in terms of intake rate) if they had. Thirdly, it might be questioned whether the birds show sampling at all. Although there is much individual variation, I conclude that they do sample because they do better than random choice, but do not show fixation on either disc, (except for bird one in treatment $b=1$).

There are several difficulties with these conclusions. Most importantly, there is a great deal of individual variation. This is most notable in the $b=1$ treatment for bird one, in which the bird seemed not to sample; this is true despite the fact that bird one clearly showed sampling in the $b=4$ treatment. There are two reasons to

think that this treatment may be a mistake, or the result of insufficient control. Firstly, this treatment was performed about one and one half months after the first two treatments. Secondly, this treatment was the last treatment, which might have caused a confusion effect, i.e. after three changes of parameters the birds might be expected to be less sure that the constant side is really constant. This problem could be resolved if I had been able to finish the third bird for all treatments. In general, more individuals must be tested before I can make more positive statements about the significance of the observed individual variation. Further experiments are presently underway. It is not uncommon, in tests of optimal foraging models, to find that average behaviour is predicated more accurately than individual behaviour (e.g. Waddington & Holden 1979, Krebs et al. 1978). I conclude at the molar level of analysis that the predictions of the theory are well supported. However, when performance is examined the birds do not as well as they might, and in the molecular analysis it can be seen that the birds do not obey the predicted rule very well at all. This departure, in molecular detail, is enough to account for the poor performance of the birds.

Implications for the theory. Why should the birds be so insensitive to environmental change? As I mentioned in chapter eight, there are two dubious assumptions, immediate recognition and that change is a first order Markov process. Violations of either of these assumptions will tend to increase the cost of sampling and hence favour less sampling. Both of these assumptions are probably unnatural for great tits. It is possible that these assumptions are so artificial that the birds simply could not learn to take advantage of them. These

results point toward further work in two directions. Theoretically, we need to know more about optimal behaviour in a changing environment where there is imperfect recognition and more complicated rules governing environmental change. However, these are difficult mathematical problems. It seems likely that even if they should be solved, their solutions may be too complicated to be of much use to behavioural ecologists. A more productive approach may be to seek simple and natural foraging systems where the mathematics can be built around the natural history rather than the other way around. For example, it may be possible to find foraging systems where immediate recognition is a perfectly valid and natural assumption (perhaps Estabrook and Jespersen's 1974 model-mimic systems). My point is that we may be able to make stronger inferences about complex systems, if we "dissect" simple and tractable systems. In section 10.6 I discuss such a naturally analytical problem in another context.

Summary In this chapter I present experimental evidence on the effect of altering the ratio of the cost of sampling errors to the cost of over-run errors (b) on the sampling behaviour of great tits. I find that sampling changes in the predicted way, but that the absolute amount of sampling is too small. The birds are more insensitive to environmental change than predicted. There are several difficulties with the interpretation of the results, most importantly individual variation in behaviour. However, this experiment is important as the first experimental test of sampling in a changing environment.

CHAPTER TEN

The Value of Patch Sampling

10.1 Introduction

In this chapter I consider the problem of sampling patches, although the ideas developed may be more generally applicable. I consider sampling in patch use because it is an area of broad ecological interest and controversy (Charnov 1976, Oaten 1977, Green 1980, Iwasa et al 1981, McNamara 1982). One difficulty in understanding the controversy surrounding patch sampling is that all of the models have made different assumptions; for example, Charnov (1976) assumes continuous energy gain and patch depression, while Green's (1980) model, which he uses to criticize Charnov, assumes that energy is gained in discrete lumps at a constant non-decreasing rate with ⁱⁿ each patch. To what extent do these differences make comparisons between models dubious? In this chapter I ask this question analytically. I do this by defining the value of patch sampling (section 10.2) and using this definition to examine whether such things as the presence of recognizable alternatives (section 10.4) or patch depression (section 10.5) can be changed independently of the value of sampling.

I also consider two other aspects of the value of sampling. In section 10.3 I examine the effect of contagion (or spread) in the mixing distribution of patch sub-types being sampled (that is, how different are the patch sub-types ^{between which} the forager cannot distinguish?),

on both the value of sampling and the value of optimal behaviour with respect to certain types of rules of thumb. In the first five sections I criticize the approach of conventional models of patch sampling; however I acknowledge that patch sampling may often be

important. In section 10.6 I consider a fairly natural example of patch sampling, and show that it has support from field observations.

10.2 A definition of the value of sampling

Optimal sampling behaviour is often complex. For example, Oaten's (1977) sampling model of patch use assumes that the forager somehow stores a vector of data consisting of the time and value of each prey capture; in contrast, Charnov's (1976) "perfect patch recognition" model only requires that the forager account for the time since entering the patch. In some sense, complicated rules for behaviour are probably more costly than simple rules. This has lead several authors to speculate that animals ought to apply simple rules of thumb in sampling situations (cf. Cowie and Krebs 1979, Janetos and Cole 1981 and Houston et al in press). One motivation for considering the value of sampling is to find a number such that if it is high, then it will be worthwhile to implement complex and expensive sampling rules, but when it is low, it may be better to adopt simpler and cheaper procedures.

The Definition: I restrict my attention to sequentially encountered patches, and the assumption of rate maximization. Imagine that the forager can recognize (without cost) k different types (or classes) of patches. Call each class C_i , $i \in \{1, \dots, k\}$. The class C_i is itself composed of n_i initially indistinguishable sub-types, $(s_{i,1}, s_{i,2}, \dots, s_{i,j}, \dots, s_{i,n_i})$. If the forager encounters a patch of class C_i it only knows that it has encountered one of the n_i patch sub-types making up C_i .

For simplicity, I consider a case in which decisions are made at intervals (decision intervals). Associated with each state $s_{i,j}$ is an integer $m_{i,j}$ (which may be infinite) which is the maximum number

of decision intervals possible if $s_{i,j}$ obtains. During the w^{th} decision interval the forager observes a random variable Q_w , which may be a vector, (Q_w might be the number of prey seen or captured, or the size of prey captured in interval w , etc.). A sampling strategy is any strategy which depends on the realizations of Q_h , $h \in \{1, \dots, w-1\}$; where $w-1$ represents the interval just past. Conversely, a strategy is not a sampling strategy if it does not depend on the values of the Q_h 's. This definition is similar to Houston et al.'s (in press) closed-loop (sampling) and open-loop (non-sampling) strategies.

Let U be the universal set of all possible strategies, including sampling and non-sampling strategies. Let u be a particular member of U . Let $e(C_i, u)$ be the expected net energy gain per patch from patches of type (class) C_i given that strategy u is employed. Similarly, let $T(C_i, u)$ be the expected time per patch. The rate of energy gain is given by

$$R = \frac{\sum_{i=1}^k p_i e(C_i, u_i)}{\tau + \sum_{i=1}^k p_i T(E_i, u_i)} \quad (10.1)$$

where τ is the travel time between patches, p_i the proportion of patch encounters with the class C_i (because this is the fine-grained case I assume that the p_i 's are not functions of the u_i 's), and the u_i are strategies adopted in patches of type C_i . Let $(u_1^*, u_2^*, \dots, u_k^*)$ be the strategies chosen from U to maximize R .

Now consider an arbitrary subset of U called Ω . In general, Ω is a set of rules of thumb which may be defined by some common attribute.

ω is a particular member of Ω . Exactly as above we define $(\omega_1^*, \omega_2^*, \dots, \omega_k^*)$ to be strategies chosen from Ω which maximize R as given by (10.4).

The value of optimal behaviour with respect to the set of rules Ω is $W(\Omega)$, where W stands for worth:

$$W(\Omega) = \frac{\sum_{i=1}^k p_i e(C_i, u_i^*)}{\tau + \sum_{i=1}^k p_i T(E_i, u_i^*)} - \frac{\sum_{i=1}^k p_i e(C_i, \omega_i^*)}{\tau + \sum_{i=1}^k p_i T(C_i, \omega_i^*)} \quad (10.2)$$

In particular we are interested in the set of strategies Ω_N which is the set of all non-sampling strategies. I define $W(\Omega_N)$ to be the value of sampling, and symbolically let $W_S = W(\Omega_N)$.

In itself, this is a fairly hollow piece of symbolism. Except as an aid to making our ideas about sampling and rules of thumb more concrete it is not very useful. It can be made more interesting by fleshing it out with specific assumptions of the sets of U and Ω_N and the classes of patches (the C_i 's). In the next three sections I consider the value of sampling and the value of optimal behaviour with respect to certain rules of thumb in specific cases.

10.3 The self moderation problem.

In chapters eight and nine I studied a simple sampling problem. One theoretical conclusion was that if the gains due to a sampling strategy were compared to a non-sampling "specialist" strategies, the conditions for sampling to be valuable (i.e. the best alternative) were very limited. I also suggested, in chapter eight, that this may be a fairly general phenomenon. Figure 10.1 illustrates this argument. Imagine that the forager cannot distinguish two states of nature represented by the two overlapping normal distributions in the figure,

Figure 10.1 Large and Small Overlaps.

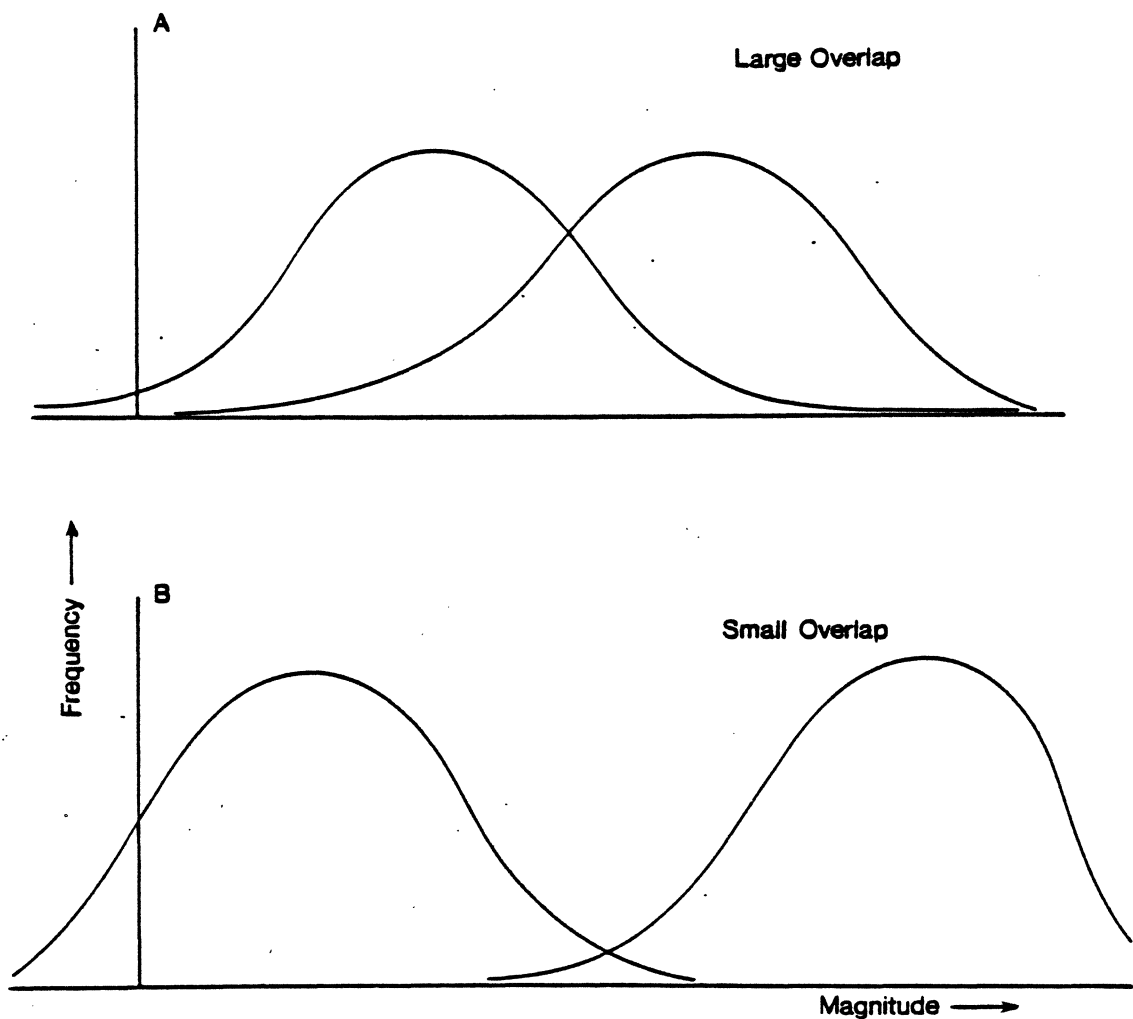


Figure 10.1 The figure shows two pairs of overlapping normal distributions. The Self-moderation Problem suggests that when overlap is large (A) , distinguishing between two sub-types (distributions) will be difficult but not important; but when over-lap is small (B) distinguishing will be more important, but also easier.