J. theor. Biol. (1980) 85, 673–690

The Application of Statistical Decision Theory to Animal Behaviour

JOHN McNAMARA†

Department of Statistics, University of Sussex, England

AND ALASDAIR HOUSTON

Animal Behaviour Research Group, Department of Zoology, University of Oxford

(*Received* 9 *November* 1979, *and in revised form* 21 *February* 1980)

Statistical decision theory is discussed as a general framework for analysing how animals should learn. Attention is focused on optimal foraging behaviour in stochastic environments. We emphasise the distinction between the mathematical procedure that can be used to find optimal solutions and the mechanism an animal might use to implement such solutions. The mechanisms might be specific to a restricted class of problems and produce suboptimal behaviour when faced with problems outside this class. We illustrate this point by an example based on what is known in the literature on animal learning as the partial reinforcement effect.

1. Introduction

Although the application of optimality principles to biology has been criticised, we agree with the justification offered by Maynard Smith (1978). As he points out "... in testing a model we are not testing the general proposition that nature optimizes but the specific hypotheses about constraints, optimization criteria and heredity" (p. 35). In accordance with this view, we do not claim that animals are optimal statistical decision makers. Instead we try to make clear what the optimal solution to a statistical problem involves. Our aim is to develop a conceptual framework for discussing the optimality of behaviour under conditions of uncertainty. Although this framework is completely general, we confine ourselves to a discussion of feeding behaviour. Our reason is that this area presents many well-defined problems and has frequently been investigated using optimality principles. These investigations have resulted in what is known as optimal foraging theory (reviewed by Pyke, Pulliam & Charnov, 1977; Krebs, 1978). Most of this work has been deterministic, and has been criticised on this

† Present address: School of Mathematics, University of Bristol, England.

0022-5193/80/160673+18$02.00/0

ground by Oaten (1977). Although his paper makes an important contribution to this area, Oaten does not provide a systematic account of stochastic problems. We attempt to give such an account in section 2. The method described involves modifying one's estimate of the probability of various possible states of nature ("models") being true. This emphasizes that the solution to stochastic problems depends on the set of models considered. In section 3 we discuss the difficulties that arise when the experimenter and the animal under study have different models in mind.

Our use of the phrase "in mind" with respect to the animal is not meant to imply that it consciously considers a class of models, or even that it reaches the optimal solution by procedures that correspond to those of section 2. It is much more likely that animals have simple "rules-of-thumb" that may approximate the optimal solution. The analysis of a problem in terms of optimality says nothing about the form such rules may take. This means that when an animal is faced with a problem that it has not been "designed" to solve, its decision rules may result in behaviour which cannot be directly understood by the application of optimality principles. In section 4 we illustrate this idea by considering the problem of how long an animal should persist in a behaviour that no longer results in food being obtained. We derive the policy that maximizes the expected energetic gains and discuss the way in which a decision rule could achieve this behaviour. Attention is drawn to the possibility of these rules producing some "odd" decisions when the animal is faced with other problems.

2. Statistical Decision Problems

In this section we consider the mathematical formulation of optimal sequential statistical decision problems. Abstract definitions will be illustrated by the following two examples. Although real animals are faced with similar problems, we do not wish to suggest that they solve them by the methods outlined in this section.

Example 1. In an experiment to examine foraging strategies an animal is placed in a cage. At one end of this cage is an apparatus which can dispense food items. The animal can obtain food from this apparatus by pressing a lever. When the lever is pressed the apparatus always gives a small food item whose size is always the same: we shall call it a half unit of food. At the other end of the cage is another machine with its own lever. However, this apparatus only gives a reward with probability θ when its lever is pressed. The reward is always one unit of food. We suppose that the animal does not know θ, but knows that it remains constant throughout the experiment.

We shall refer to the act of pressing one of the two levers as a trial. The apparatus is arranged so that after every trial, the animal must return to the centre of the cage before it can execute another trial. The animal is then faced with a decision problem known as a one-armed bandit problem. The "one" refers to the fact that only one machine's performance is unknown. The machine which always gives a reward $\frac{1}{2}$ is, in this case, the known arm of the bandit, and the machine which gives a reward with probability θ is called the unknown arm. [In the decision problem discussed by Krebs *et al.* (1978) the probabilities of success, p_1 and p_2, on the two arms are both unknown. This decision problem is therefore known as a two-armed bandit problem.]

Example 2. We give the following idealised model of a kingfisher catching food. The environment consists of identical pools of water. Each pool either has one fish in it or is empty. If there is a fish present, the times at which the fish appears on the surface of the pool form a Poisson process with rate 1. The kingfisher catches fish by waiting by a pool until a fish appears and then catching it. The bird knows in advance that a proportion γ of pools have a fish in them. The time to travel from one pool to the next is a constant, τ.

A decision problem may involve a stochastically varying system without the problem being classified as a statistical decision problem. For example, suppose in example 1 that the expected food intake over 20 trials is to be maximized. If θ is known the problem is completely trivial: if $\theta > \frac{1}{2}$ trials are always made on the stochastic apparatus, if $\theta < \frac{1}{2}$ trials are always made on the apparatus which gives a reward of $\frac{1}{2}$ unit. In this case, although the outcome of a trial on one apparatus is probabilistic, the probability distribution governing the outcome of this trial is known in advance: the results of trials give no further information about the underlying system and the optimal policy can ignore their outcome. When θ is not known the situation is completely different. In this case the results of trials give information about the true value of θ and the optimal policy will be truly sequential, taking the outcome of previous trials into account at each stage.

This illustrates that in a statistical decision problem, not only are the outcomes of observations or actions probabilistic in nature, but the probability distributions are themselves unknown.

For a given statistical decision problem, the true state of nature will be characterized by a particular value of a parameter θ, although this value will not usually be known. We denote the set of possible values of θ by Ω. For each θ in Ω (written $\theta \in \Omega$) there is a corresponding probability model E_θ. The set of all such models will be called ξ: that is

$$\xi = (E_\theta : \theta \in \Omega)$$

(the collection of all models E_θ as θ ranges over Ω). Thus in example 1, without further information restricting the range of θ we have a collection of probability models

$$\xi_1 = (E_\theta\colon 0 < \theta < 1),$$

where one member of this collection describes the actual performance of the unknown arm.

Consider a particular pool in example 2. There are only two probability models for this pool. We have

$$\xi_2 = (E_i\colon i = 0, 1),$$

where in model E_0 there is no fish in the pool and in model E_1 there is one fish in the pool. Let T be the waiting time till a fish appears and $P_i(A)$ denote the probability of event A under model E_i. The behaviour of the system under model E_0 is given by

$$P_0(T \leq t) = 0 \quad \text{for all } t \geq 0,$$

and the behaviour under E_1 by

$$P_1(T \leq t) = 1 - \mathrm{e}^{-t} \quad \text{for all } t \geq 0.$$

(A) PRIOR DISTRIBUTIONS

Given a family of probability models $\xi = (E_\theta\colon \theta \in \Omega)$, one must have some prior knowledge about the likelihood that any particular E_θ represents the true state of nature. Without such knowledge it is not possible to talk about maximizing expected rewards and therefore impossible to talk about optimality. We use example 2 to illustrate this. Consider the following policy: if a fish appears before time t_0 eat it and immediately move on to the next patch; if no fish has appeared before time t_0 move on to the next patch at this time. It is meaningless to ask what the average feeding rate is from this policy unless the proportion of pools with fish is known in advance. The average rate depends on γ, the proportion of pools of type E_1, and hence the optimal time to wait at a given pool depends on γ. Thus the term "optimal policy" is not well defined unless γ is specified in advance. This type of argument is quite general.

Let $\xi = (E_\theta\colon \theta \in \Omega)$ be the family of possible probability models for the system in question. We assume a prior distribution π is given for this family: that is we assume for each $\theta \in \Omega$ there is a prior probability $\pi(\theta)$ that E_θ represents the true state of nature.

Of course, just because a prior probability is needed to formulate the mathematical problem, this does not mean that priors are well defined in nature, nor that an animal uses them. However, in an environment that is not changing rapidly over time it is valid to ascribe probabilities to various states and hence to use priors. We discuss this point further in section 3.

For example 2 we have

$$\pi(0) = \text{Probability pool of type } E_0 = 1 - \gamma,$$

$$\pi(1) = \text{Probability pool of type } E_1 = \gamma.$$

We have assumed that pools are indistinguishable in this example. If this is not the case then $\pi(1)$ need not be γ. In a more realistic situation the prior would not just be based on the relative frequency of pool type E_1 to E_0, but on information such as past experience of this particular pool, the appearance and size of the pool, the time of day, weather conditions, etc. It will be assumed that a prior can be based on any relevant information available to a foraging animal.

In example 1, let us first consider the simplified version with family of models ξ_1. The prior will be taken to be a probability density $\pi(\theta)$ on $0 \leq \theta \leq 1$. It is mathematically convenient to take π to be a β-distribution with parameters α and β. That is

$$\pi(\theta) \propto \theta^{\alpha-1}(1-\theta)^{\beta-1} \qquad 0 \leqslant \theta \leqslant 1 \tag{1}$$ †

The prior probability for the success on the unknown arm is then $(\alpha/\alpha + \beta)$.

The prior on a more realistic family of models would be suitably complex.

(B) OBSERVATION AND THE POSTERIOR DISTRIBUTION

Let us consider the simplified version of example 1 with family of models ξ_1. Let the random variable X be the amount of food obtained from a trial on the unknown arm: that is $X = 0$ if the trial is a failure and $X = 1$ if it is a success. Let $P_\theta(X = x)$ be the probability that $X = x$ when E_θ describes the true state of nature. We have

$$P_\theta(X = x) = \begin{cases} 1-\theta & \text{if } x = 0 \\ \theta & \text{if } x = 1. \end{cases} \tag{2}$$

We can generalize this to n trials as follows. Let X_i be the amount of food obtained on the ith trial on the unknown arm, $i = 1, \ldots, n$. Let $\mathbf{X}$ denote the vector $(X_1, \ldots, X_n)$ and $\mathbf{x}$ denote the vector $(x_1, \ldots, x_n)$. We have

$$P_\theta(\mathbf{X} = \mathbf{x}) = \theta^r(1-\theta)^{n-r} \tag{3}$$

where $r = \sum_{i=1}^n x_i$ is the number of successes on the n trials.

† The constant of proportionality is chosen so that $\int_0^1 \pi(\theta)\, d\theta = 1$.

Now consider example 2 with family of models ξ_2. Let T be the time to catch a fish in a particular pool. After time t, an observation consists in either the value of T if a fish has been caught, or the observation $T > t$ if no fish has been caught. Let $P_\theta(T > t)$ be the probability that $T > t$ under model E_θ. We have

$$\begin{aligned} P_0(T > t) &= 1, \\ P_1(T > t) &= e^{-t}. \end{aligned} \tag{4}$$

In both cases information about the values of certain random variables is obtained during foraging; and for a given observation, $\mathbf{X} = \mathbf{x}$ in example 1 and $T > t$ or $T < t$ in example 2, there is a corresponding function of θ, as given in equations (3) and (4) respectively.

In general we have a family of probability models $\xi = (E_\theta\colon \theta \in \Omega)$. Suppose that a set of observations (or single observation) is taken. We can describe the outcome of these observations by a random variable X. Let $P_\theta(\mathbf{X} = \mathbf{x})$ be the probability that $\mathbf{X} = \mathbf{x}$ when θ is the true state of nature. Thus, after observing the value of $\mathbf{X}$ as $\mathbf{x}$ we have a function

$$P_\theta(\mathbf{X} = \mathbf{x}) \qquad \theta \in \Omega$$

which we consider as a function of θ for fixed $\mathbf{x}$. This is called the likelihood function for the observation $\mathbf{X} = \mathbf{x}$.

Initially we have a prior distribution $\pi(\theta)$ for θ. Subsequent observations give additional information about which model, E_θ, is the true model for the system. To be more precise the observation $\mathbf{X} = \mathbf{x}$ allows us to deduce a new value $\pi^*(\theta) \equiv \pi(\theta|\mathbf{x})$, for the probability that E_θ describes the true state of nature. This revised probability distribution is called the posterior distribution for θ given $\mathbf{X} = \mathbf{x}$. It is given by Bayes rule

$$\pi^*(\theta) \equiv \pi(\theta|\mathbf{x}) = \frac{\pi(\theta) P_\theta(\mathbf{X} = \mathbf{x})}{\sum\limits_{\alpha \in \Omega} \pi(\alpha) P_\alpha(\mathbf{X} = \mathbf{x})}. \tag{5}$$

Note that the posterior is proportional to the prior times the likelihood function. We illustrate this formula below.

Let us consider example 1 with the class of models ξ_1 and prior distribution given by equation (1). Suppose n trials have been made on the unknown arm. By equation (1) the prior distribution is proportioned to $\theta^{\alpha-1}(1-\theta)^{\beta-1}$, and by equation (3) the likelihood function is $\theta^r(1-\theta)^{n-r}$ where r is the number of successes. Thus the posterior distribution is proportional to $\theta^{\alpha+r-1}(1-\theta)^{\beta+n-r-1}$: which tells us that the posterior is also a β-distribution with parameters α_n and β_n where

$$\begin{aligned}\alpha_n &= \alpha + r \\ \beta_n &= \beta + n - r.\end{aligned} \tag{6}$$

In example 2, if a fish has been observed then it is known for sure that the pool is of type E_1. So we consider the case where no fish has been observed by time t. Equations (4) and (5) give the posterior probability that there is a fish in the pool, given no fish has been observed by time t, as

$$\frac{\gamma\,\mathrm{e}^{-t}}{1+\gamma(\mathrm{e}^{-t}-1)}.$$

We now consider the case of independent observations. For simplicity we assume that observations are taken sequentially at discrete "times" $t = 1, 2, \ldots$, as in example 1. Let Y_n be the random variable observed at time n. We make the assumptions that the variables Y_1, $Y_2, \ldots$ are independent for every probability model θ. This assumption is very often met in practice and is obeyed by observations in example 1 and 2 and by observations in the partial reinforcement model discussed in section 4. At time $t = 0$ the prior $\pi(\theta)$ contains all the available information about the true value of θ. If we observe $Y_1 = y_1$ at time $t = 1$, then the posterior $\pi^1(\theta) \equiv \pi(\theta|y_1)$ may be calculated using equation (5). This function now contains all the available information about the true value of θ. Once it has been determined the original prior π and the first observation y_1 are redundant and can be forgotten. Thus $\pi^1(\theta)$ acts as a new prior distribution for θ. Another observation, $Y_2 = y_2$ can be taken and the posterior $\pi^2(\theta)$ calculated using equation (5)

$$\pi^2(\theta) \equiv \pi^1(\theta|y_2) = \frac{\pi^1(\theta)P_\theta(Y_2 = y_2)}{\sum_{\alpha\in\Omega} \pi^1(\alpha)P_\alpha(Y_2 = y_2)}$$

and this in turn acts as a new prior, and so on.

(C) THE OPTIMAL POLICY—ACTIONS, REWARDS AND OPTIMALITY CRITERIA

At any given time and for any given state there will be a set of behaviours open to an animal. A foraging animal might continue searching in a given patch, move to a new patch or hide from predators. More generally, an animal may display to a prospective mate, build a nest, migrate, etc. We will refer to the set of behaviours open to an animal at a given time as the set of actions; this may depend on the state of the animal and its environment at that time. The action taken at time t will be denoted by a_t. As a result of this action the animal may have additional information available to it in the form of an observation, $Y_t = y_t$. For simplicity we assume that actions are taken at discrete times $t = 0, 1, 2, \ldots$ and that the random variables $Y_0, Y_1, Y_2 \ldots$ are independent.

In example 1, if the animal is only concerned with food intake the only relevant actions we need consider are:

(a) make a trial on the known arm,
(b) make a trial on the unknown arm and
(c) rest.

We can associate a reward with each state and action taken in that state. In foraging problems the reward is often taken to be the expected amount of food consumed in the next unit of time. Thus in example 1 action (a) has reward $\frac{1}{2}$, the reward for action (b) is the current posterior probability of success on the unknown arm and action (c) has reward zero. The reward can be considerably more general than the amount of food eaten. Energy costs may be taken into account. The animal may not wish to consume as much food as possible but maximize the probability that it has sufficient food to eat during the course of a day. There may be predators about, in which case the reward for a food item may be its food value minus a cost due to the associated risk of predation.

In general, if we are considering optimal behaviour in animals, the reward for an action must ultimately be related to the contribution of that action towards fitness. We do not intend to discuss this relationship in this paper, but take the reward for an action as a given quantity. Unfortunately these quantities are not usually known with any precision in practice. Discussions of the contributions of actions to fitness may be found in Krebs (1978) or Maynard Smith (1978).

To complete the characterisation of the problem we have to specify what function of the rewards is to be maximised. The three most commonly used criteria for stochastic problems are to maximize:

(i) The expected reward over some time interval of duration T. T is called the time horizon for the problem.
(ii) The average reward per unit time.
(iii) The expected discounted reward: where rewards incurred at time t in the future are discounted by a factor $e^{-\phi t}$.

In summary, an optimality problem consists of a class of probability models $\xi = (E_\theta : \theta \in \Omega)$ and a prior distribution $\pi(\theta)$, together with a set of actions and rewards and an optimality criterion.

Details of how optimal policies are derived can be found in De Groot (1970) or Ross (1970). The optimal solution consists in specifying what action should be taken in any circumstance. In assessing any action, we have to consider the expected reward from the action, the resulting change in the animal's state (e.g. increases in information) and the value of being in the new state. The rewards and the change of state depend on the current posterior distribution, while the value of the new state depends on the optimal policy in the future. This policy will, of course, depend on all the components mentioned above. In particular, if criterion (i) is used the optimal solution depends on the time horizon.

Some of these points can be illustrated by considering example 1. We take ξ to be the family ξ_1 of probability models with prior distributions given by equation (1). Let the two possible actions be (a) make a trial on the known arm and (b) make a trial on the unknown arm. Let the reward on any trial be the amount of food eaten and the optimality criterion be to maximize the total number of rewards obtained (i.e. amount of food eaten) over a specified number, T, of trials. In this problem the "time" $t = 0, 1, 2, \ldots$ is the number of trials that have been made. Note that trials on the known arm give no information about θ and hence if action a_n is (a) then $\pi^{n+1}(\theta) = \pi^n(\theta)$. The posterior after n trials is thus a β-distribution with parameters $\alpha_{n'}$ and $\beta_{n'}$ [see equation (6)] where n' is the number of trials that have been conducted on the unknown arm. It is clear that the optimal action at time n must be determined solely by $\alpha_{n'}$, $\beta_{n'}$ and n.

These parameters determine the current probability of a success on the unknown arm. When there are only a few trials to go trials are made on the unknown arm if this probability is greater than $\frac{1}{2}$ and on the known arm if this quantity is significantly less than $\frac{1}{2}$. When there are still many trials to go trials may be made on the unknown arm even when current probability of success is significantly less than $\frac{1}{2}$. However the situation is not that simple since a decision is never made to continue indefinitely on the unknown arm, only to continue provided β_n is larger than some critical value.

It should be emphasised that the optimal policy is strongly dependent on the number of trials, T, which are available. Trials on the unknown arm serve a dual purpose in giving a possible reward and providing information about the true value of θ. When few trials are available it is more important to gain rewards than gather information and there is a tendency to make trials on the arm which appears better. When many trials are available initially they tend to be made on the unknown arm to gain information. The policy of immediate maximization, that is of always choosing the arm that appears best, is equivalent to the optimal policy for time horizon $T = 1$. Since the optimal policy varies with time horizon the policy of immediate maximization will not be optimal in general.

3. What Problems is the Animal Built to Solve?

In the previous section we described what is known as the Bayesian approach to statistical decision problems. Fundamental to this approach is the assignment of a prior distribution to the class of possible states of nature. If the environment is stable, both the class of models and the prior distribution will be well-defined. The action of natural selection on a species that has evolved in such an environment will tend to make the behaviour of a member of this species close to the optimal solution for the problem posed by this environment. This could be brought about by direct genetic control, or the animal could be built to learn the relevant parameters. This does not mean that the decision process used by the animal involves a class of models with a prior distribution. We believe it is more reasonable to assume that the animal has relatively simple decision rules ("rules of thumb"). If these rules are approximately optimal, they must reflect the structure of the decision problem, and hence the resulting behaviour looks as if the animal was using a class of models and a prior probability distribution. We will call this class the class to which the animal is adapted. Species will differ in the class of models to which they are adapted, and it is important to work with the correct class in considering any given species.

We can illustrate this point with reference to example 1. In this case the experimenter knows that the probability of success on the unknown arm is a constant, θ, and that the reward on the known arm is always a half unit. It is likely, however, that an animal exposed to this experimental situation will be adapted to a wider class of models. In nature the probability of success or local average rate of return for some action is rarely a constant over space or time. For example, foraging patches tend to be depleted, so that an animal will receive a diminishing return if it stays in a patch. Thus an animal faced with the experimental set-up of example 1 will be adapted to a class that contains not only the models ξ_1, but also probability models where the parameter θ changes, and in particular where it decreases over time. This means that the animal will have difficulty in deciding whether statistical fluctuations in the number of successes on the unknown arm are due to θ changing during the course of an experiment, or due to chance. In fact there is evidence that an animal interprets such fluctuations as a change in θ, as we discuss below.

A simple thought experiment suggests that animals allow for the possibility of such changes. Suppose in example 1 that the optimality criterion is that the animal should maximize its expected food intake over a day (10 h). We assume that the animal has been exposed to this experimental situation several times before and is fairly sure that the reward on the known arm is always $\frac{1}{2}$. Suppose initially that the animal always succeeds on the unknown arm. If ξ_1 is taken as the class of probability models, then, for any reasonable prior, an optimizing animal would settle down to feeding entirely from the unknown arm after a short period of time. This would probably also be the case in practice. Let us suppose that the animal feeds from this arm for the first 5 h of the day. Then halfway through the day the unknown arm stops giving rewards (and remains that way for the rest of the day). How long should the animal remain on the unknown arm? If the family of probability models is restricted to ξ_1, then an optimizing animal would continue on the unknown arm until the posterior mean on this arm falls to close to $\frac{1}{2}$. For any reasonable prior this would take the rest of the day. An animal which obeyed this policy would have five hours of failure with the known arm always available but unused. This is not a strategy that a real animal would indulge in. The policy of continuing on the unknown arm for the rest of the day seems absurd to us because we too have evolved and lived in a variable and changing environment; and so regard a long run of failures as evidence that θ has changed, regardless of what we have been told to the contrary. We conclude that the animal recognizes that it is a possibility that resources may run out; that is it behaves as if it considers models with decreasing θ as possible probability models for the system.

In the above thought experiment the length of time an animal should continue on the unknown arm depends on the likelihood of various changes in θ; that is on the prior and consequent posterior probability of various models in which θ changes. This prior is presumably determined partly by the environment in which the animal's ancestors evolved and partly by the animal's previous experience. We discuss a similar experimental situation, involving a giving-up time, in section 4.

The above arguments raise a problem for experiments proporting to test the optimality of behaviour. An animal's behaviour in an experiment depends on both its previous experience of the world and the environment in which its ancestors evolved. Even if it behaves optimally with respect to the class to which it is adapted, this class is unlikely to coincide with the experimenter's class of models, and hence its policy will not coincide with the theoretical optimal policy derived by the experimenter. As an example of this problem we cite the experiment of Krebs, Kacelnik & Taylor (1978). In this two-armed bandit experiment there was a tendency for a great tit to give up on a particular arm after a run of failures, when the (experimenter's) optimal policy dictated that the animal should continue on this arm. This suggests that the bird "interpreted" the run of failures as an indication that the probability of success had decreased, rather than as a statistical fluctuation, as was in fact the case.

4. Partial Reinforcement—When to Give Up

In this section we consider a problem derived from experiments on what is known as the partial reinforcement effects (PRE). The basic finding is is that if animals are always rewarded for making a response ("continuous reinforcement") they stop responding sooner in subsequent extinction (a period when no responses are rewarded) than animals that were only rewarded on a certain percentage of responses ("partial reinforcement"). A clear review of the literature on this subject can be found in Mackintosh (1974).

Although the PRE has sometimes been mentioned in the context of optimal foraging (see, for example, Smith & Dawkins, 1971) there has been no attempt at a rigorous derivation of the optimal policy. In the first part of this section we obtain the optimal solution for the general problem of how long to persist when responses no longer yield rewards. We then consider a mechanism which could produce this solution.

Consider an animal that can make a response which has a certain probability of producing a food reward. If the animal has a run of unrewarded responses, this could be a statistical fluctuation but might mean that the response was no longer effective. We apply Bayesian decision theory to the problem of deciding how many unrewarded responses should be made if the energetic returns are to be maximized.

We model the problem by assuming that the probability of success on a given trial (i.e. of reward following a given response) is a known constant p. Extinction is represented by assuming that there is some unknown number N, such that the probability of success is p on the first N trials, but all subsequent trials are failures. If there have been at least N trials we will say the resource has run out.

Let E_i be the probability model in which an animal receives a reward with probability p on the first i trials, but receives no reward on subsequent trials. We take as collection of probability models the set $\xi = (E_i: i = 0, 1, 2, \ldots)$, and assume some prior distribution π_i, $i = 0, 1, 2, \ldots$, where π_i is the prior probability that $N = i$.

Suppose after n trials there have been m successes, and that the jth success occurred on the k_jth trial, where $1 \leq k_1 < k_2 \ldots < k_m$. It can easily be verified [using equation (5)] that the posterior distribution after n trials is a function of k_m and n alone. Thus the present time and time of last success summarize all the information obtained over the first n trials: further details of the pattern of successes are superfluous, not even m need be known. (Of course, in practice, the value of m is relevant to the animal's estimation of p.)

We first consider prior distributions of the form

$$\pi_i = (1 - d)d^i \qquad i = 0, 1, 2, \ldots, \tag{7}$$

where $0<d<1$. One can interpret d by noting that $d/1-d$ is the mean value of N and hence is the expected number of trials the resource will last. The tabulations below consider $d = 0{\cdot}9$ and $d = 0{\cdot}975$. These values lead to average trial numbers of 9 and 39 respectively.

If the nth trial is a success then immediately afterwards the posterior probability that $N = j$ is zero if $j = 0, 1, 2, \ldots, n-1$, and $(1-d)d^{j-n}$ if $j > n$. Thus, after a success, the probability that the resource will run out after a further i trials is just $(1-d)d^i$, which is the original prior. Therefore when the prior has the form of equation (7) we do not even need to remember the number of trials, just the number of trials since the last success.

To illustrate the amount of information given by a run of failures and its dependence on p, we tabulate the probability that resources have not run out against the number of failures since the last success.

TABLE 1

Probability that resources have not run out

Number of failures	0	1	2	3	4	5	6	7
$p = 0{\cdot}5$	0·90	0·74	0·52	0·32	0·17	0·084	0·049	0·018
$p = 0{\cdot}9$	0·90	0·43	0·062	0·0059	$<10^{-3}$	$<10^{-4}$	~0	~0
$p = 0{\cdot}99$	0·90	0·074	$<10^{-3}$	~0	~0	~0	~0	~0

$d = 0{\cdot}9$

TABLE 2

Probability that resources have not run out

Number of failures	0	1	2	3	4	5	6	7
$p = 0{\cdot}5$	0·98	0·93	0·84	0·71	0·54	0·36	0·21	0·12
$p = 0{\cdot}9$	0·98	0·76	0·25	0·032	0·0032	$<10^{-3}$	$<10^{-4}$	~0
$p = 0{\cdot}99$	0·98	0·27	0·0037	$<10^{-4}$	~0	~0	~0	~0

$d = 0{\cdot}975$

Here ~0 indicates a quantity smaller than 10^{-5}. These tables show that for a given number of failures, the nearer p is to unity the more sure we are that the run of failures is not a statistical fluctuation and that the resource has run out.

To give an idea of whether it is worth giving up after a certain number of failures we plot the total number of future successes we expect if trials are continued indefinitely. If resources had not run out this number would increase with p, but as Tables 3 and 4 show this is more than offset by the increasing probability that resources have run out.

TABLE 3

Expected number of future successes

Number of failures	0	1	2	3	4	5	6	7
$p = 0{\cdot}5$	4·5	3·7	2·6	1·6	0·86	0·42	0·20	0·09
$p = 0{\cdot}9$	8·1	3·9	0·56	0·053	0·005	$<10^{-3}$	$<10^{-4}$	~ 0
$p = 0{\cdot}99$	8·9	0·79	0·0071	$<10^{-4}$	~ 0	~ 0	~ 0	~ 0

$d = 0{\cdot}9$

TABLE 4

Expected number of future successes

Number of failures	0	1	2	3	4	5	6	7
$p = 0{\cdot}5$	19·5	18·6	16·9	14·2	10·6	7·2	4·3	2·3
$p = 0{\cdot}9$	35·1	27·9	9·0	1·1	0·12	0·01	$<10^{-3}$	$<10^{-4}$
$p = 0{\cdot}99$	38·6	10·8	0·14	0·0014	$<10^{-4}$	~ 0	~ 0	~ 0

$d = 0{\cdot}975$

A specification of rewards and optimality criterion is necessary before anything further can be said about the optimal time to stop. We will assume that the animal has only two actions available at a given time: to continue with the next trial or to give up. Let the reward on a success be one unit and suppose a trial costs 0·1 units of energy. Finally, let the optimality criterion be to maximize the expected total reward minus costs. With this model the number of failures before giving up is shown in Table 5.

TABLE 5

Number of failures before giving up

$d =$	0·9	0·975
$p = 0{\cdot}5$	6	11
$p = 0{\cdot}9$	3	4
$p = 0{\cdot}99$	2	3

Thus, we see that the closer p is to unity, the sooner an optimizing animal would stop, which is in agreement with the findings of Weinstock (1958).

Although we have considered a particular form of reward and optimality criterion, it is the behaviour under varying p shown in Tables 1 and 2 which is crucial. We would obtain qualitatively similar results for a slightly different choice of rewards and optimality criterion, such as a reward for giving up.

We have also obtained these results for a specific form of prior [equation (7)]. For an arbitrary prior the number of failures before giving up will be a function of the total number of trials as well as p. However, it will still generally be true that the giving up time is a decreasing function of p for a given number of trials. A more realistic form of prior than equation (7) might be one which tended to zero more slowly as i tended to infinity. With such a prior the giving up time is an increasing function of the number of trials for fixed p. Thus, the longer trials have been going on and the smaller p (provided it is not very small), the larger the giving up time.

These results can be criticized because the class, ξ, of probability models is too simple. In nature p will not be known exactly, nor will it be constant. Also resources do not end abruptly as we have envisaged. The same type of criticism was made of the work of Krebs, Kacelnik & Taylor (1978) earlier in this paper. We believe, however, that these simple models are useful provided their limitations are realised. For example, it is not realistic to extend the above analysis to the case $p = 1$. In this case the optimal policy is to give up immediately a failure occurs. However, when $p = 1$ it becomes important that p is not known exactly and that resources do not end abruptly, so that we should be wary of making quantitative predictions on the basis of the above analysis.

The conclusion to be drawn from this discussion is that, regardless of the exact details, the optimal policy for this sort of problem involves persisting for more trials in the face of failures if p is low. This provides an explanation of the PRE in terms of optimality theory.

Our analysis of this problem says nothing about how an animal might achieve the optimal solution. It is possible, though unlikely, that the mechanism corresponds to the mathematical procedures outlined in section 2.A. It seems more plausible to assume that the animal has some simple rule of thumb. Such rules may not produce optimal behaviour in circumstances to which the animal is not adapted, as the following example shows.

We consider a mechanism which has two systems, one sensitive to rewards and the other sensitive to punishments (the latter including the failure to get an expected reward). The mechanism continues to make responses as long as the output R of the reward system is greater than the output P of the punishment system. The value of R is proportional to the magnitude of the reward the mechanism "expects" and hence is constant for a given reward size. On each unrewarded trial there is an input F (for failure) to the punishment system. This input is proportional to the magnitude of the expected reward (i.e. F is proportional to R). A run of unrewarded trials results in a weighted sum F_s of the F inputs. This sum is multiplied by a gain factor k to give P. The value of k determines the sensitivity of the mechanism to unrewarded trials. When k is low F_s has to be large before $F_sk = P$ exceeds R, and this requires a long run of failures. When a reward occurs F_s is reset to zero and the value of k is reduced by an amount that is an increasing function of F_s. Because F_s increases with the length of a run of failures and such runs will be long when p is low, it can be seen that k will be small when p is low. In this way the mechanism "learns" to respond in the face of failure. [Cf. "Partial reinforcement ensures that subjects are reinforced for making the required ... response in a situation similar to that normally encountered in extinction." Mackintosh (1974)] Extinction corresponds to giving the mechanism nothing but failures. When k is low, many failures are required before P becomes greater than R, so the mechanism makes more responses in extinction when p is small than when it is large, as is required by the optimization argument. In particular, if the mechanism has never experienced non-reward (i.e. has received continuous reinforcement), its value of k will not have been reduced. As a consequence it will require fewer failures before it stops responding than when it has had unrewarded trials (i.e. partial reinforcement), which corresponds to the basic PRE.

Other parallels between the mechanism and experimental results can be drawn. For example, under conditions of continuous reinforcement, the mechanism will stop sooner during extinction if the preceding rewards were large, because this produces a bigger value of F. This corresponds to the findings of Hulse (1958). We observed that if an animal is sure that $p = 1$, then it should stop after the first failure. If we extend our theory to include things like a gradual decrease in p, which would make some persistence reasonable, this persistence should increase rather than decrease with reward size. When animals are given partial reinforcement, this increase is indeed found (Hulse, 1958; Likely, Little & Mackintosh, 1971), which may indicate that animals are not adapted to continuous reinforcement. Similarly, our proposed mechanism approximates the optimal solution for a given problem but produces "paradoxical" behaviour when confronted with problems for which it was not designed. A simple example of the "side-effects" of a decision rule is provided by the response of an incubating gull when faced with a choice between its own egg and a model egg that is much bigger than a real egg (a "super-normal" stimulus). A gull's choice of the large model can be interpreted in terms of a rule for choosing the largest gull egg available. Under natural conditions the constraints on egg size make this a reasonable rule, but by presenting the gull with eggs outside its experience, bizarre results can be obtained.

As a second example, it can be seen that giving the mechanism a long run of rewards after a period of partial reinforcement will not result in it stopping sooner during extinction. The reason for this is that k is not reduced by the session of continuous reinforcement because it only gets changed when a run of unrewarded trials is followed by a reward. The behaviour of the mechanism is in agreement with the results obtained by Sutherland, Mackintosh & Wolfe (1965) from rats.

This example is important because it emphasises the difference between our approach and the discrimination hypothesis (Mowrer & Jones, 1945), which claims that when p is low it is hard to know that extinction has begun. This hypothesis cannot account for the existence of the PRE if partial reinforcement is followed by continuous reinforcement (Jenkins, 1962; Sutherland, Mackintosh & Wolfe, 1965). By separating the optimality analysis from the mechanism, we do not have to predict that the PRE would be abolished in this case. The reason for this is that the mechanism does not "estimate" p by averaging over all trials, but by the length of sequences of unrewarded trials. As a result, paradoxical effects occur when long bouts of continuous reinforcement are given. This suggests that the mechanism (and the animal) is not adapted to such circumstances.

It must be emphasised that we are making no special claims for the mechanism we have discussed in this section. Our choice was dictated by convenience rather than by consideration of the exact nature of reward and punishment mechanisms. Furthermore, many aspects of the mechanism have not been specified, so that its response to certain problems is undefined. We do not, however, think that this weakens our case. Our aim was to illustrate a statistical decision problem and emphasise that animals might solve it by mechanisms that differ from the ones a mathematician would use. The data on the PRE illustrate this point and suggest that many of the results reported may have no direct interpretation as optimal strategies. They may be no more than side-effects of rules which approximate the optimal solution in other contexts.

We thank John Krebs, Alex Kacelnik, John Bather, Stephen Lea and an anonymous referee for helpful comments on the manuscript. J.Mc. was supported by a SCR fellowship, A.I.H. was supported by a SCR grant to David McFarland and John Krebs.

REFERENCES

DEGROOT, M. H. (1970). *Optimal Statistical Decisions.* New York: McGraw-Hill.

HULSE, S. H., Jr. (1958). *J. exp. Psychol.* **56,** 48.

JENKINS, H. M. (1962). *J. exp. Psychol.* **64,** 441.

KREBS, J. R. (1978). In: (J. R. Krebs & N. B. Davies, eds), *Behavioural Ecology.* Oxford: Blackwell Scientific Publications.

KREBS, J. R., KACELNIK, A. & TAYLOR, P. (1978). *Nature* **275,** 27.

LIKELY, D., LITTLE, L. & MACKINTOSH, N. J. (1971). *Can. J. Psychol.* **25,** 130.

MACKINTOSH, N. J. (1974). *The Psychology of Animal Learning*. London: Academic Press.
MAYNARD SMITH, J. (1978). *Ann. Rev. Ecol. Syst.* **9,** 31.
MOWRER, O. H. & JONES, H. M. (1945). *J. exp. Psychol.* **35,** 293.
OATEN, A. (1977). *Theor. Popul. Biol.* **12,** 263.
PYKE, G. H., PULLIAM, H. R. & CHARNOV, E. L. (1977). *Quart. Rev. Biol.* **52,** 137.
ROSS, S. M. (1970). *Applied Probability Models with Optimisation Applications*. San Francisco Holden-Day.
SMITH, J. N. M. & DAWKINS, R. (1971). *Anim. Behav.* **19,** 695.
SUTHERLAND, N. S., MACKINTOSH, N. J. & WOLFE, J. B. (1965). *J. exp. Psychol.* **69,** 56.
WEINSTOCK, S. (1958). *J. exp. Psychol.* **46,** 151.