# Optimal Foraging and Learning 

John M. McNamara<br>School of Mathematics, University of Bristol<br>AND<br>Alasdair I. Houston<br>Department of Zoology, University of Oxford

(Received 9 May 1985)
Optimal foraging theory usually assumes that certain key environmental parameters are known to a foraging animal, and predicts the animal's behaviour under this assumption. However, an animal entering a new environment has incomplete knowledge of these parameters. If the predictions of optimal foraging theory are to hold the animal must use a behavioural rule which both learns the parameters and optimally exploits what it has learnt. In most circumstances it is not obvious that there exists any simple rule which has both these properties. We consider an environment composed of well-defined patches of food, with each patch giving a smooth decelerating flow of food (Charnov, 1976). We present a simple rule which (asymptotically) learns about and optimally exploits this environment. We also show the rule can be modified to cope with a changing environment. We discuss what is meant by optimal behaviour in an unknown and possibly changing environment, using the simple rule we have presented for illustrative purposes.

## Introduction

Optimal foraging theory is based on the assumption that natural selection will favour the foraging strategy that maximizes the forager's fitness (see Krebs et al., 1983; Pyke et al., 1977; Pyke, 1984 for reviews). What Krebs et al. (1983) call classical optimal foraging theory assumes that fitness is maximized by the maximization of the net rate of energetic gain. There are theoretical reasons why this net rate is not necessarily the appropriate currency (Caraco, 1980; McNamara \& Houston, 1982), and there is evidence that considerations other than net rate determine foraging decisions (e.g. Caraco et al., 1980; Caraco, 1983; Caraco \& Lima, 1985). Nevertheless, we will work within the classical optimal foraging framework. The advantage of doing so is that it is easier to relate our arguments to previous work in this area. It is also worth pointing out that (a) net rate may sometimes be
an appropriate currency and (b) the issues that we raise remain relevant whatever the currency for foraging may be.
The first models of optimal foraging assumed that the forager "knew" the parameters of the environment (e.g. Charnov, 1973; Pulliam, 1974). Oaten (1977) pointed out that rewards could provide a forager with information about its environment, but the forager was assumed to have knowledge of the possible environmental states. The forager is using information but not learning anything new about the environment as a whole. It can be argued that Krebs et al. (1978) were the first people to raise the issue of optimal learning. Great tits (Parus major) were given a series of choices between two feeding sites. Each site had a constant probability of giving the bird a food item if the bird chose it. At the start of a test, the reward probabilities were not known to the bird. This procedure corresponds to the two-armed bandit problem of decision theory. Problems such as this are difficult because they involve a combination of parameter estimation and the maximization of payoff. The optimal policy can be found by Bayesian decision theory in which the animal can be thought of as forming estimates of the reward probabilities (McNamara \& Houston, 1980). There is, however, no need to assume that animals actually use such procedures. Simple rules can perform almost as well as a Bayesian decision maker (Houston et al., 1982).
Houston et al. (1982) consider simple learning rules within a framework based on optimization. In contrast to this approach, Ollason (1980) sees his learning rule as an alternative to an optimality analysis. Furthermore, he argues ( p .51 ) that if an animal is learning then it cannot be foraging optimally and if it is foraging optimally it cannot be learning. Ollason applied his learning rule to an environment of depleting patches. His simulations suggest that when there are several patch types his learning rule approaches the optimal behaviour. We use the same paradigm to explore the following issues: (a) is it always possible to learn the classical optimal strategy, i.e. the strategy which maximizes long term reward rate? (b) What does it mean to forage optimally in an unknown, and possible fluctuating, environment?

## The Model Environment

We consider the patchy environment first analyzed by Charnov (1973, 1976). We assume that there are $k$ patch types labelled $E_{1}, E_{2}, \ldots, E_{k}$. The proportion of patches of type $E_{i}$ is $\alpha_{i}$, so that the probability that the next patch visited is type $E_{i}$ is always $\alpha_{i}$. Patches are not revisited. On a patch of type $E_{i}$ rewards are gained as a smooth flow: the rate after time $t$ has
elapsed on a patch being $r_{i}(t)$. We assume that, for each $i, r_{i}(t)$ is a strictly decreasing function which tends to zero as $t$ tends to $\infty$. The travel time between patches, $\tau$, is a random variable with a finite expectation $0<E(\tau)<\infty$.

We assume that, on encountering this environment, the animal does not know $k, \alpha_{i}, r_{i}$ or $\tau$.

The marginal value theorem (Charnov, 1976) says that the long term rate is maximized if the forager leaves each patch when the rate $r_{i}(t)$ drops to the maximum possible long-term rate. This rate will be denoted by $\gamma^{*}$. The theorem provides a rule that specifies the optimal behaviour if $\gamma^{*}$ is known. We are assuming, however, that the environmental parameters, and hence $\gamma^{*}$, are not known. The forager is now faced with the problem that optimal behaviour (in its simplest sense) requires a knowledge of $\gamma^{*}$, but $\gamma^{*}$ can only be achieved by behaving optimally. Despite this apparent circularity, we show in this section that it is always possible to learn $\gamma^{*}$ in the environment that we are considering.

A general way in which behaviour can be represented is to view the state of the animal (or model animal) as determining behaviour. The consequences of behaviour in turn modify the state and hence determine future behaviour (e.g. Houston et al., 1977). Within this framework, a rule can be said to learn about an environmental parameter when the value of a state variable converges to the parameter value in question. In the case of patch-use, the parameter that we will consider is the maximum possible rate $\gamma^{*}$. We now describe a simple rule that learns the value of $\gamma^{*}$.

## The Learning Rule

The animal encounters patches sequentially. Let

$$
\begin{aligned}
& t_{n}=\text { time spent on the } n \text {th patch } \\
& \tau_{n}=\text { time spent travelling between the } n \text {th and } n+1 \text { th patch. } \\
& T_{n}=t_{n}+\tau_{n} .
\end{aligned}
$$

Thus $T_{1}+\ldots+T_{n}$ is the time between arrival at the first patch and arrival at the $n+1$ st. Let

$$
R_{n}=\text { reward obtained on the } n \text {th patch. }
$$

We consider a rule in which $\gamma^{*}$ is recursively estimated. Initially constants $R_{0}$ and $T_{0}$ are chosen ( $R_{0}, T_{0}>0$ ). The ratio $\gamma_{0}=R_{0} / T_{0}$ is the initial estimate for $\gamma^{*}$. The animal leaves the first patch encountered when the reward rate on this patch falls to $\gamma_{0}$. On arrival at the next patch it forms a new estimate
$\gamma_{1}$ for $\gamma^{*}$, and uses this to determine when to leave the second patch, and so on. We consider a rule which has the following detailed form.
(a) On arrival at the $n+1$ th patch $(n=0,1, \ldots)$ the animal forms the estimate $\gamma_{n}$ for $\gamma^{*}$ where

$$
\begin{equation*}
\gamma_{n}=\frac{R_{0}+R_{1}+\ldots+R_{n}}{T_{0}+T_{1}+\ldots+T_{n}} \tag{1}
\end{equation*}
$$

(b) The animal leaves the $n+1$ th patch when the reward rate on this patch falls to $\gamma_{n}$.
To illustrate the properties of this learning rule, we first consider a simple environment in which all patches are the same. In this case it is possible to give a reasonably easy proof of results which are true in the more complex environments that we subsequently consider.

## Learning in the Simplest Case

Consider an environment in which there is only one patch type with reward rate $r(t)$, and the travel time $\tau$ is a constant.


Fig. 1. A graphical illustration of the relationship between an estimate $\gamma$ for $\gamma^{*}$ and the long term reward rate $f(\gamma)$ which results from using this estimate on each patch encountered. The solid curve is the cumulative reward $R(t)=\int_{0}^{t} r(v) d v$ as a function of time $t$ on patch. For any estimate $\gamma$, the rule is to leave each patch when $r(t)$ falls to $\gamma$. The resulting rate $f(\gamma)$ is the cumulative reward $R(t)$ divided by the total time $(t+\tau)$. Three values of $\gamma$ are illustrated ( $\gamma_{1}, \gamma^{*}$ and $\gamma_{2}$ ). In each case the rate is the slope of the solid straight line that joins the curve to the point $Q$. When $\gamma=\gamma^{*}$, then $f(\gamma)=\gamma^{*}$. In the other two cases $\gamma$ is indicated by the slope of the broken line.

Suppose that the animal uses the same estimate $\gamma>0$ for $\gamma^{*}$ on every patch it encounters. Let $f(\gamma)$ be the long term reward rate achieved by using this estimate. Figure 1 illustrates the dependence of $f(\gamma)$ on $\gamma$. It can be seen that
(a) If $\gamma=\gamma^{*}$ then the animal leaves at time $t^{*}$ and achieves rate $f\left(\gamma^{*}\right)=$ $\gamma^{*}$.
(b) If $\gamma=\gamma_{1}$ where $0<\gamma_{1}<\gamma^{*}$, the animal leaves at time $t_{1}$ and achieves rate $f\left(\gamma_{1}\right)$ where $\gamma_{1}<f\left(\gamma_{1}\right)<\gamma^{*}$.
(c) If $\gamma=\gamma_{2}$ where $\gamma_{2}>\gamma^{*}$ then the animal leaves at time $t_{2}$ and achieves rate $f\left(\gamma_{2}\right)$ where $f\left(\gamma_{2}\right)<\gamma^{*}$.
Note that if $\gamma>r(0)$ then each patch is left immediately and hence $f(\gamma)=0$. Figure 2 shows the full function $f$ when $r(t)=\mathrm{e}^{-t}$ and $\tau=1$. In this case $\gamma^{*}=0.31784$ and the optimal time on patch is $t^{*}=1 \cdot 1462$. As can be seen from the figure, conditions (a), (b) and (c) are all satisfied in this case.


Fig. 2. The relationship between an estimate $\gamma$ of $\gamma^{*}$ and the long term reward rate $f(\gamma)$ which results from using this estimate on each patch encountered. The curve gives $f(\gamma)$ when $r(t)=\mathrm{e}^{-t}$ and $\tau=1 \cdot 0$. The points of intersection of this curve with the straight line give solutions to the equation $f(\gamma)=\gamma$. Note that $\gamma=\gamma^{*}$ is the only positive solution to this equation.

The graphical results, illustrated in Fig. 1, are true in general and are established by analytical argument in Appendix 1. We summarise the main conclusions.

$$
\begin{align*}
& 0 \leq f(\gamma) \leq \gamma^{*} \text { for all } \gamma>0  \tag{2}\\
& f\left(\gamma^{*}\right)=\gamma^{*}  \tag{3}\\
& 0<\gamma<\gamma^{*} \Rightarrow \gamma<f(\gamma) . \tag{4}
\end{align*}
$$

In particular $\gamma^{*}$ is the only positive root of the equation $f(\gamma)=\gamma$ : i.e.

$$
\begin{equation*}
f(\gamma)=\gamma \quad \text { for } \gamma>0 \Rightarrow \gamma=\gamma^{*} \tag{5}
\end{equation*}
$$

We now apply these results to our learning rule. Equations (2), (3) and (4) show that the estimates $\gamma_{n}$ and $\gamma_{n+1}$ satisfy the following relationships

$$
\begin{align*}
& \gamma_{n}>\gamma^{*} \Rightarrow 0<\gamma_{n+1}<\gamma_{n}  \tag{6}\\
& \gamma_{n}=\gamma^{*} \Rightarrow \gamma_{n+1}=\gamma^{*}  \tag{7}\\
& 0<\gamma_{n}<\gamma^{*} \Rightarrow \gamma_{n}<\gamma_{n+1}<\gamma^{*} \tag{8}
\end{align*}
$$

These relationships are proved in Appendix 2. They imply the following result.

$$
\begin{equation*}
\gamma_{n} \rightarrow \gamma^{*} \text { as } n \rightarrow \infty \tag{9}
\end{equation*}
$$

The proof can be divided into two parts.
Part 1. Proof that $\gamma_{n}$ tends to a limit
Suppose that $0<\gamma_{0}<\gamma^{*}$. Then by equation (8)

$$
\gamma_{0}<\gamma_{1}<\gamma_{2}<\ldots<\gamma^{*}
$$

Thus $\gamma_{n}$ tends to a limit as $n$ tends to infinity.
Suppose $\gamma_{0}=\gamma^{*}$. Then, by equation (7), $\gamma_{n}=\gamma^{*}$ for all $n$.
Finally, suppose that $\gamma_{0}>\gamma^{*}$. Then, by equation (6), one of the following two possibilities must occur
(a) $\gamma_{0} \geq \gamma_{1} \geq \gamma_{2} \geq \ldots \geq \gamma^{*}$, or
(b) $0<\gamma_{N}<\gamma^{*}$ for some $N$.

If (a) occurs then $\gamma_{n}$ must tend to a limit as $n$ tends to infinity. If (b) occurs then by equation (8)

$$
\gamma_{N}<\gamma_{N+1}<\gamma_{N+2}<\ldots<\gamma^{*}
$$

and again we have convergence.
In conclusion $\gamma_{n}$ tends to a limit as $n$ tends to infinity. We denote this limit by $\tilde{\gamma}$. It can be seen from the above proof that $\tilde{\gamma}>0$.

Part 2. Proof that $\tilde{\gamma}=\gamma^{*}$
Since $\gamma_{n}$ tends to a limit so do the rewards $R_{n}$ and times $T_{n}$. We denote these limits by $R$ and $T$ respectively. By equation (1) $\gamma_{n}$ must tend to $R / T$, so that $R=\tilde{\gamma} T$.

It can be shown that $f(\gamma)$ is a continuous function of $\gamma$ for $\gamma>0$. Thus $f\left(\gamma_{n}\right)$ tends to $f(\tilde{\gamma})$ as $n$ tends to infinity. But $R_{n+1}=f\left(\gamma_{n}\right) T_{n+1}$, since $f\left(\gamma_{n}\right)$ is the average rate the animal would obtain by using the estimate $\gamma_{n}$ on each patch. Thus taking the limit as $n$ tends to infinity we find that $R=f(\tilde{\gamma}) T$. Comparing this with the equation $R=\tilde{\gamma} T$ we conclude that $f(\tilde{\gamma})=\tilde{\gamma}$. Thus from condition (5) it can be seen that $\tilde{\gamma}=\gamma^{*}$.


Fig. 3. The convergence of $\gamma_{n}$ to $\gamma^{*}$ when $r(t)=\mathrm{e}^{-t}$ and $\tau=1 \cdot 0$. In this case $\gamma^{*}=0.31784$. In (a) $\gamma_{0}=0.6>\gamma^{*}$. The triangles correspond to $T_{0}=2, R_{0}=1.2$, the squares to $T_{0}=0.75$, $R_{0}=0.45$ and the circles to $T_{0}=0.01, R_{0}=0.006 . \operatorname{In}(b) \gamma_{0}=0.15<\gamma^{*}$. The squares correspond to $T_{0}=2.5, R_{0}=0.375$, the triangles to $T_{0}=0.8, R_{0}=0.12$ and the circles to $T_{0}=0.01, R_{0}=$ $0 \cdot 0015$.

The convergence of $\gamma_{n}$ to $\gamma^{*}$ is illustrated in Fig. 3, from which the effect of variations in the choice of $R_{0}$ and $T_{0}$ can be seen. In Fig. 3(a) $\gamma_{0}$ is greater than $\gamma^{*}$ and in Fig. 3(b) $\gamma_{0}$ is less than $\gamma^{*}$. When $R_{0}=0.006$, $T_{0}=0.01$, it can be seen that $\gamma_{0}>\gamma^{*}$ and $\gamma_{1}<\gamma^{*}$. In all the other examples illustrated, convergence is monotone.

## Learning in the General Case

When there is more than one patch type present (or when $\tau$ varies) $T_{n}$ and $R_{n}$ are random variables and the sequence $\gamma_{0}, \gamma_{1}, \gamma_{2}, \ldots$ is a stochastic process. We illustrate this with three stochastic environments which are all simple modifications of the environment in which each patch has reward rate $r(t)=\mathrm{e}^{-t}$.

In each environment there are two patch types $E_{1}$ and $E_{2}$, with parameters $k_{1}$ and $k_{2}$ respectively. On a patch with parameter $k_{i}(i=1,2)$ we have
$r_{i}(t)=k_{i} \exp (-t)$. The two types are present in equal proportions so that $\alpha_{1}=\alpha_{2}=0 \cdot 5$. The travel time $\tau$ is again equal to 1 . The three environments are described in Table 1 together with the optimal stay times on each patch type. The table also gives the resulting long term average rate on each patch type (rewards on patch divided by time on patch plus $\tau$ ). For each environment parameters have been chosen so that the maximum long term average rate, $\gamma^{*}$, is again equal to $0 \cdot 31784$.

Table 1

| Environment | Patch type | $k$ | $t^{*}$ | Rate |
| :---: | :---: | :--- | :---: | :---: |
| A | $E_{1}$ | 0.8 | 0.9231 | 0.2507 |
|  | $E_{2}$ | 1.1823 | 1.3137 | 0.3736 |
| B | $E_{1}$ | 0.6 | 0.6354 | 0.1725 |
|  | $E_{2}$ | 1.3277 | 1.4297 | 0.4156 |
| C | $E_{1}$ | 0.4 | 0.2299 | 0.0668 |
|  | $E_{2}$ | 1.4203 | 1.4971 | 0.4415 |

To illustrate the stochastic nature of the process, consider Environment $C$ and suppose that the initial constants are

$$
T_{0}=1 \quad \text { and } \quad R_{0}=0.31784 \equiv \gamma^{*},
$$

so that $\gamma_{0}=\gamma^{*}$. The possible values of $\gamma_{1}, \gamma_{2}$ and $\gamma_{3}$ are given in Fig. 4. As can be seen there are two possible values of $\gamma_{1}$, four possible values of $\gamma_{2}$, and eight possible values of $\gamma_{3}$. Each time that a good patch (type $E_{2}$ ) is encountered, $\gamma_{n}$ increases; each time that a poor patch is encountered, $\gamma_{n}$ decreases. Because $\alpha_{1}=\alpha_{2}=0 \cdot 5$, the eight possible values of $\gamma_{3}$ are equally likely.


Fig. 4. Possible values of $\gamma_{1}, \gamma_{2}$ and $\gamma_{3}$ for Environment C when $T_{0}=1$ and $R_{0}=0.31784 \equiv$ $\gamma^{*}$.

It is important to note that the order in which patch types are visited influences $\gamma_{n}$. For example the sequence $E_{2}, E_{2}, E_{1}$ results in $\gamma_{3}=0 \cdot 3606$, whereas the sequence $E_{1}, E_{2}, E_{2}$ results in $\gamma_{3}=0.3518$. The reason for this effect is that the patch types that have been visited determine the current estimate of $\gamma$, which in turn determines the time spent on the current patch. Thus the time spent on a patch type and the rewards obtained from it are not fixed, but depend on previous experience. As a result, even when $\gamma_{0}=\gamma^{*}$ it is not clear that $\gamma_{n}$ tends to $\gamma^{*}$ as $n$ tends to infinity. The difficulty is that even though the proportion of patches of type $i$ that are visited tends to $\alpha_{i}$ as the number of patches visited tends to infinity, the order in which patches are visited is also relevant. It is therefore not possible to extend the proof given above by simply averaging. By approaching the problem in a different way, it is possible however to prove that equation (9) holds with probability one. This result follows from an application of a general result on Markov renewal processes given in McNamara (1985). Strictly speaking the general theory presented there only applies to the case where each patch visited eventually runs out of food, so that $r_{i}(t)=0$ for $t$ sufficiently large. The theory can however be extended in various ways to cover other cases.

Ollason (1980) considers a rule which has some similarities to the rule we consider. One major difference is that the past is discounted by his rule. The degree of discounting is measured by a parameter $k$ which tends to infinity as memory stretches further back into the past. He considers the special case of rewards of the form $r(t)=a e^{-b r}$ in detail. In this case his rule has the property that, when all patch are identical, the time spent on patch tends to a limit, $t_{c}$, as time in the environment tends to infinity. As $k$ tends to infinity $t_{c}$ tends to the optimal time on patch. When patches differ his simulations suggest that behaviour tends towards optimal as time in the environment and $k$ tend to infinity.

This suggests that setting $k=\infty$ might give a rule which, like our rule, is asymptotically optimal (see below). As can be seen from his equation (1) this is not the case. Setting $k=\infty$ here leads to a rule under which a patch is never left until $r(t)=0$. Such a rule would perform very badly.

## Rates of Convergence

Clearly it is advantageous to obtain accurate estimates for $\gamma^{*}$ as quickly as possible. To measure how quickly the estimate $\gamma_{n}$ tends to $\gamma^{*}$ we introduce the error $\sigma_{n}^{2}$ defined by

$$
\sigma_{n}^{2}=E\left\{\left(\gamma_{n}-\gamma^{*}\right)^{2}\right\} ;
$$

i.e. $\sigma_{n}^{2}$ is the mean square error in the estimate $\gamma_{n}$ of $\gamma^{*}$. We use $\sigma_{n}^{2}$ to
measure estimation error because the loss in reward rate that results from using the wrong $\gamma$ to determine when to leave a patch is approximately proportional to $\left(\gamma-\gamma^{*}\right)^{2}$ for small $\left|\gamma-\gamma^{*}\right|$. A justification of this claim is given in Appendix 3.

If $\gamma_{0}=\gamma^{*}$ then $\sigma_{n}^{2}=0$ for all $n$ in a non-stochastic environment. Thus in order to investigate the effect of stochasticity on $\sigma_{n}^{2}$, we set $T_{0}=1$ and $R_{0}=\gamma^{*}$, so that $\gamma_{0}=\gamma^{*}$. The resulting mean square error in the three environments $\mathrm{A}, \mathrm{B}$ and C is illustrated in Fig. 5 , which shows that the mean square error increases with increasing variability in the environment.


Fig. 5. The effect of environmental variability on the error with which $\gamma^{*}$ is estimated. The root mean square error $\sigma_{n}$ is plotted as a percentage of $\gamma^{*}$ for (from top to bottom) the three environments $\mathrm{C}, \mathrm{B}$ and A as a function of the number, $n$, of patches visited. The environments are described in Table 1, which shows that $C$ is the most variable and $A$ the least variable. The figure shows that the error increases with increasing environmental variability.

To illustrate the additional effect of a variable travel time, some values of the percentage root mean square error under variable $\tau$ are given in Table 2. The values come from Environment B, modified so that $\tau$ has mean 1 but a uniform distribution over an interval centred on 1 . The table shows that the mean square error increases with increasing variability in $\tau$.

Table 2
Learning $\gamma^{*}$ when the travel time is variable (Environment B)

| $\%$ rms error | $n=2$ | $n=4$ | $n=8$ | $n=20$ | $n=50$ |
| :--- | :---: | :---: | :---: | :---: | :---: |
| $\tau=1$ | $21 \cdot 2$ | $16 \cdot 5$ | $12 \cdot 3$ | $8 \cdot 1$ | $5 \cdot 2$ |
| $\tau \sim U\left(\frac{1}{2}, \frac{3}{2}\right)$ | $22 \cdot 7$ | 17.8 | $13 \cdot 2$ | $8 \cdot 6$ | $5 \cdot 5$ |
| $\tau \sim U(0,2)$ | 27.6 | 21.5 | $15 \cdot 8$ | $10 \cdot 2$ | $6 \cdot 5$ |

These results illustrate a phenomenon which we might expect to hold generally for any rule which estimates $\gamma^{*}$ : the more variability in the environment, the slower an estimate tends to converge to $\gamma^{*}$.

It is clear from Fig. 3 that the rate at which $\gamma_{n}$ tends to $\gamma^{*}$ also depends on $R_{0}$ and $T_{0}$. Mean square error is reduced if the initial estimate $\gamma_{0}$ is close to $\gamma^{*}$. However, it is not always possible to ensure this since $\gamma^{*}$ is $a$ priori unknown. For given $\gamma_{0}=R_{0} / T_{0}$ Fig. 3 illustrates that $\sigma_{n}^{2}$ also depends on the magnitudes of $R_{0}$ and $T_{0}$. When all patches are the same small $R_{0}$ and $T_{0}$ ensure rapid convergence. (Although, as Fig. 3 illustrates, if $\gamma_{0}>\gamma^{*}$ and $R_{0}$ and $T_{0}$ are very small $\gamma_{1}$ may overshoot $\gamma^{*}$.) However, if patches differ, this is no longer true. As Fig. 4 illustrates random fluctuations can change the estimate of $\gamma^{*}$ dramatically. The smaller the values of $R_{0}$ and $T_{0}$ the more sensitive a rule is to these fluctuations. As a consequence if one is fairly sure that the initial estimate $\gamma_{0}$ is close to $\gamma^{*}$ it is best to choose $R_{0}$ and $T_{0}$ large so that estimates will be insensitive to short term fluctuations.

It can thus be seen that the initial estimates $R_{0}$ and $T_{0}$ together define a sort of prior mean and variance for the unknown parameter $\gamma^{*}$. The ratio $\gamma_{0}=R_{0} / T_{0}$ acts as the prior mean, while the magnitude of the parameters $R_{0}$ and $T_{0}$ play the role of prior variance, with variance decreasing with increasing $R_{0}$ and $T_{0}$.

## Optimal Learning

The discussion of convergence has led us to the central issue of optimality in the context of learning. As we have already mentioned, Ollason (1980) argues that learning is incompatible with optimal foraging, and vice versa. While it is clearly true that an animal (or rule) that has to learn about the environment cannot always maximize reward rate, it does not seem reasonable to conclude that a rule for learning cannot be optimal in some broader sense. One such sense is that of asymptotic optimality, which has already been described in the context of learning by Houston et al. (1982). A rule is asymptotically optimal if its performance tends to the optimal performance as experience tends to infinity. In other words, the rule always learns the optimal policy in the long run. This criterion is more or less the one used by Harley (1981). The results described above show that there exists an asymptotically optimal rule for a broad class of patchy environments.

It can be objected, however, that asymptotic optimality is not necessarily a very good criterion, in that it assumes that performance will be assessed over an unlimited period. Rules that are asymptotically optimal may perform badly over limited periods (Houston et al., 1982). The appropriate currency depends on the details of an animal's environment. If the animal has a
fixed period in a given environment, then the total expected reward may be the best criterion. If the animal has a constant probability of being forced to stop foraging, then some form of discounted rewards should be the criterion (see Houston et al., 1982).

Establishing the optimality criterion does not, however, suffice to enable the optimal learning rule to be specified. This can be seen from our discussion of how our rule depends on $R_{0}$ and $T_{0}$, where it was seen that the best choice of these parameters depends on the type of environment likely to be encountered. In general one needs to specify the range of environments which are possible, and to specify the probability of encountering an environment of any given type. In other words one needs to specify some prior distribution on the set of possible environments.

Finally, one needs to specify what information a rule is allowed to use.

## Coping with a Changing Environment

So far, we have confined our attention to a constant, but possibly stochastic, environment. The principle advantage of learning is, however, the fact that it enables an animal to respond to changes in the environment. Many people have constructed learning rules that are based on some exponentially weighted average of past experience (Harley, 1981; Killeen, 1982; Lester, 1984; Ollason, 1980; see Kacelnik et al., in press, for a review). The basic idea of exponential weighting is that more importance is given to the recent as opposed to the remote past. We now consider a modified rule that involves exponential weighting, and investigate how well this rule performs in both changing and non-changing environments. This illustrates the conflicting pressures on a learning rule and leads us to a consideration of what it means to be optimal in a changing invironment.

Following the notation used above, $R_{n}$ is the reward obtained on the $n$th patch and $T_{n}$ is the time on the $n$th patch plus the travel time to the next patch. Like our previous rule, the rule forms an estimate of $\gamma^{*}$ using previous rewards and times and leaves a patch when the rate on the patch falls to the current estimate of $\gamma$. The previous rule formed the following estimate of $\gamma_{n}$

$$
\gamma_{n}=\frac{R_{n}+\ldots+R_{0}}{T_{n}+\ldots+T_{0}}
$$

In this estimate, all previous rewards and times have equal weight. The rule that we now consider is based on the following estimate of $\gamma_{n}$

$$
\gamma_{n}=\frac{R_{n}+\mathrm{e}^{-\alpha T_{n}} R_{n-1}+\mathrm{e}^{-\alpha\left(T_{n}+T_{n-1}\right)} R_{n-2}+\ldots+\mathrm{e}^{-\alpha\left(T_{n}+\ldots+T_{1}\right)} R_{0}}{T_{n}+\mathrm{e}^{-\alpha T_{n}} T_{n-1}+\mathrm{e}^{-\alpha\left(T_{n}+T_{n-1}\right)} T_{n-2}+\ldots+\mathrm{e}^{-\alpha\left(T_{n}+\ldots+T_{1}\right)} T_{0}} .
$$

In this equation, $\alpha$ is the exponential weighting factor ( $\alpha>0$ ) (the previous rule corresponds to $\alpha=0$ ). A simple way to express this rule is to call the numerator $Y_{n}$ and the denominator $X_{n}$. It can then be seen that $\gamma_{n}=Y_{n} / X_{n}$ where

$$
Y_{n}=R_{n}+\mathrm{e}^{-\alpha T_{n}} Y_{n-1}
$$

and

$$
X_{n}=T_{n}+\mathrm{e}^{-\alpha T_{n}} X_{n-1}
$$

To investigate the properties of this rule we first consider an unchanging environment in which all patches are the same and the travel time is constant. In such an environment it can again be shown that $\gamma_{n}$ tends to $\gamma^{*}$ as $n$ tends to infinity. The proof is a simple modification of the proof given for the original rule. As before one can establish equations (6), (7) and (8) and use these equations to show that $\gamma_{n}$ must converge to some limit, and one can again use equation (5) to show that this limit must be $\gamma^{*}$. Thus the modified rule is asymptotically optimal for any value of the exponential weighting $\alpha$. The rule proposed by Ollason (1980) does not have this property for any value of his weighting factor $k$.

Under this rule the quantities $Y_{n}$ and $X_{n}$ tend to equilibrium values as the number of patches visited tends to infinity. The equilibrium values depend on $\alpha$ and the environment. The rate at which equilibrium is approached depends on $\alpha$ and increases with increasing $\alpha$.

The point of using a rule based on some weighting or discounting of past experience is that a rapid response to change is possible. To illustrate the effect of the weighting factor $\alpha$, we have calculated the response of the rule to a change in patch quality. We assume that the rule has reached equilibrium in an environment in which the rate on all patches is $r(t)=k \exp (-t)$ and $k=1$. The value of $k$ then changes to a new value which is the same for all patches. Figure 6 shows how the estimate of $\gamma^{*}$ depends on $\alpha$ and on the number of patches that have been visited. When $\alpha$ is large the previous environment is forgotten quickly and convergence to the new $\gamma^{*}$ is rapid.

Figure 6 suggests that a large value of $\alpha$ is desirable, but this is not true if the environment is stochastic. When there is no exponential weighting, all patches are represented equally in the estimate of $\gamma^{*}$. The introduction of a weighting factor means that recent patches predominate. No matter how many patches have been visited the estimate $\gamma_{n}$ is subject to the influence of the patches which happen to have been visited recently. Consequently, when the environment is stochastic $\gamma_{n}$ will not tend to a limit, but will forever fluctuate. One can use the mean square error to measure the size of these fluctuations. When there is no weighting ( $\alpha=0$ ) the mean square


Fig. 6. Response of the rule to a change in the environment. Initially $r(t)=\mathrm{e}^{-t}$ for all patches and the estimate of $\gamma^{*}$ is its true value $\gamma^{*}=0.31784$. After $n=0$ the environment changes so that $r(t)=k \mathrm{e}^{-t}$ for all patches. In (a) $k=0.6$ so that the new $\gamma^{*}$ is 0.1907 . In (b) $k=1.3$ so that $\gamma^{*}=0.4132$. In each case squares refer to $\alpha=0.25\left(t_{r}=4\right)$, triangles to $\alpha=$ $0.06666\left(t_{r}=16\right)$ and circles to $\alpha=0.025\left(t_{r}=40\right) . \tau=1$ throughout.
error tends to zero, but when there is a non-zero weighting factor $\alpha$ the mean square error tends to a positive value. The dependence of this value on the response time $t_{r}=1 / \alpha$ for environments $\mathrm{A}, \mathrm{B}$, and C is shown in Fig. 7. The shorter the value of $t_{r}$, the more weight is put on the very recent past, and hence the more sensitive $\gamma_{n}$ is to environmental stochasticity. Thus mean square error is large for small $t_{r}$ and decreases as $t_{r}$ increases.

It can be seen from this discussion that there are conflicting pressures on the choice of $\alpha$. The need to respond to a change in the environment favours a small value of $\alpha$, but such a value may result in inappropriate responses to runs of good or bad luck.

To determine the best value of $\alpha$, we need to say what it means for a learning rule to be optimal in a changing environment. We have already said that in a stochastic environment it is necessary to specify a prior distribution on environments i.e. to specify the possible environments and their probability of occurrence. In a changing environment this information


FIG. 7. The long term equilibrium value of the $\%$ rms error plotted as a function of $t_{r}=1 / \alpha$ for environments $\mathrm{A}, \mathrm{B}$ and C .
is still required but in addition we must specify how frequently changes occur and the probability distribution of environments after a change. If the state of the environment is regarded as a vector-valued stochastic process, then the above requirements amount to having knowledge of the transition probabilities that govern the evolution of the stochastic process.
For the rule that we have been considering the best value of $\alpha$ depends on the probability that the environment will change. If a change is likely then a large $\alpha$, and hence small response time, are likely to be advantageous. This advantage is offset, however, by the stochasticity of the environment. The interaction between these conflicting demands is quite complex.

## Discussion

Optimal foraging theory usually assumes that certain environmental parameters are known. Ollason (1980) stressed that animals have to learn about their environment and discussed this problem in the context of patch use. The marginal value theorem (Charnov, 1976) specifies optimal behaviour in terms of the maximum possible rate $\gamma^{*}$, but as Ollason pointed out, if a predator does not know $\gamma^{*}$, then it can only experience $\gamma^{*}$ by behaving optimally. Thus it is not clear that an animal can learn to behave optimally. We have presented a simple rule and shown that the rule learns $\gamma^{*}$ under a wide range of conditions.

Ollason (1980) argues that because a learning animal does not spend all its time foraging at the maximum rate $\gamma^{*}$ it is not foraging optimally. While this is obviously true in terms of classical foraging theory, it is inappropriate to use such an optimality criterion when the environmental parameters are
unknown. We argue in this paper that the optimality criterion must take account of the fact that an animal has to learn. It is still possible to ask if an animal is foraging as well as possible, given that it starts without complete knowledge of its environment. We have described the ingredients that are needed to formulate an optimality criterion for behaviour in a constant but unknown environment and in a changing environment. Although this involves more ingredients than are required in classical foraging theory, it is still possible for optimal behaviour to be well-defined.

Our simulations suggest that it may be difficult to learn when the environment is highly variable. The basic model involves only variability in patch types; even then learning is slow. When variability of travel times is added, learning is slower. The foraging environment of most animals is presumably more variable than any of the cases that we have considered. Consequently, it may be very difficult for animals to learn enough about their environments to forage optimally in the terms of classical foraging models.
A.I.H. was supported by a Science and Engineering Research Council grant to J. M. McNamara and J. R. Krebs.

## REFERENCES

Caraco, T. (1980). Ecology 61, 119.
Caraco, T. (1983). Behav. Ecol. Sociobiol. 12, 63.
Caraco, T. \& Lima, S. L. (1985). Anim. Behav. 33, 216.
Caraco, T., Martindale, S. \& Whittam, T. S. (1980). Anim. Behav. 28, 820.
Charnov, E. L. (1973). Optimal Foraging-some theoretical considerations. Ph.D. thesis, University of Washington.
Charnov, E. L. (1976). Theor. Pop. Biol. 9, 129.
Harley, C. B. (1981). J. theor. Biol. 89, 611.
Houston, A. I., Halliday, T. R. \& McFarland, D. J. (1977). Med. Biol. Eng. Comput. 15, 49.
Houston, A. I., Kacelnik, A. \& McNamara, J. M. (1982). In: Functional Ontogeny (McFarland, D. J. ed.). London: Pitman.
Kacelnik, A., Krebs, J. R. \& Ens, B. (in press). In: The Quantitative Analysis of Behaviour, Vol. 6 (Commons, M. L., Shettleworth, S. J. \& Kacelnik, A. eds). New York: Lawrence Erlbaum.
Killeen, P. R. (1982). In: Nebraska Symposium on Motivation (Bernstein, D. J. ed.). Lincoln: University of Nebraska Press.
Krebs, J. R., Kacelnik, A. \& Taylor, P. (1978). Nature 275, 27.
Krebs, J. R., Stephens, D. W. \& Sutherland, W. J. (1983). In: Perspectives in Ornithology (Brush, A. H. \& Clark, G. A. eds). Cambridge: Cambridge University Press.
Lester, N. P. (1984). Behaviour 89, 175.
McNamara, J. M. (1985). J. Appl. Prob. 22, 324.
McNamara, J. M. \& Houston, A. I. (1980). J. theor. Biol. 85, 673.
McNamara, J. M. \& Houston, A. I. (1982). In: Functional Ontogeny (McFarland, D. J. ed.). London: Pitman.
Oaten, A. (1977). Theor. Pop. Biol. 12, 263.
Ollason, J. G. (1980). Theor. Pop. Biol. 18, 44.

Pulliam, H. R. (1974). Am. Nat. 108, 59.
Pyke, G. H. (1984). Ann. Rev. Ecol. Syst. 15, 523.
Pyke, G. H., Pulliam, H. R. \& Charnov, E. L. (1977). Quart. Rev. Biol. 52, 137.

## APPENDIX 1

## Properties of $\boldsymbol{f}(\boldsymbol{\gamma})$

Let $\gamma>0$. Since $r(t)$ is monotonic decreasing and tends to zero as $t$ tends to infinity there exists $t \equiv t(\gamma)$ such that

$$
\begin{gather*}
r(t)>\gamma \quad 0 \leq t<t(\gamma)  \tag{A1.1}\\
r(t(\gamma)) \leq \gamma . \tag{A1.2}
\end{gather*}
$$

Note that, since $r(t)$ is a decreasing function of $t, t(\gamma)$ is a decreasing function of $\gamma$.

We now define $R$ by

$$
\begin{equation*}
R(t)=\int_{0}^{t} r(v) \mathrm{d} v \quad t \geq 0 \tag{A1.3}
\end{equation*}
$$

and define $\gamma^{*}$ by

$$
\begin{equation*}
\gamma^{*}=\sup _{t \geq 0} \frac{R(t)}{(t+\tau)} \tag{A1.4}
\end{equation*}
$$

Finally, we define

$$
\begin{equation*}
f(\gamma)=\frac{R(t(\gamma))}{t(\gamma)+\tau} \quad \gamma>0 \tag{A1.5}
\end{equation*}
$$

We show that $f$ satisfies equations (2), (3) and (4).
Equation (2) holds by equations (A1.4) and (A1.5), and equation (3) is just a restatement of the Marginal Value Theorem. Thus it only remains to establish equation (4). Let $\gamma$ satisfy $0<\gamma<\gamma^{*}$. Then, since $t(\gamma)$ is monotone decreasing, we have $t(\gamma) \geq t\left(\gamma^{*}\right)$. We thus have

$$
\begin{equation*}
R(t(\gamma))=\int_{0}^{t(\gamma)} r(v) \mathrm{d} v=R\left(t\left(\gamma^{*}\right)\right)+\int_{t\left(\gamma^{*}\right)}^{t(\gamma)} r(v) \mathrm{d} v . \tag{Al.6}
\end{equation*}
$$

Now by equation (Al.5)

$$
R\left(t\left(\gamma^{*}\right)\right)=\left(t\left(\gamma^{*}\right)+\tau\right) f\left(\gamma^{*}\right)
$$

Thus by equation (3) and the assumption $\gamma<\gamma^{*}$ we have

$$
\begin{equation*}
R\left(t\left(\gamma^{*}\right)\right)>\gamma\left(t\left(\gamma^{*}\right)+\tau\right) \tag{A1.7}
\end{equation*}
$$

Also by equation (A1.1)

$$
\begin{equation*}
\int_{t\left(\gamma^{*}\right)}^{t(\gamma)} r(v) \mathrm{d} v \geq \int_{t\left(\gamma^{*}\right)}^{t(\gamma)} \gamma \mathrm{d} v=\gamma\left(t(\gamma)-t\left(\gamma^{*}\right)\right) \tag{A1.8}
\end{equation*}
$$

Thus from equations (A1.6), (A1.7) and (A1.8) we have

$$
R(t(\gamma))>\gamma(t(\gamma)+\tau)
$$

Equation (4) then follows from the definitions of $f(\gamma)$ (equation A1.5).

## APPENDIX 2

## Relationship of $\boldsymbol{\gamma}_{\boldsymbol{n}+\boldsymbol{1}}$ to $\boldsymbol{\gamma}_{\boldsymbol{n}}$

We prove equations (6), (7) and (8).
The learning rule prescribes that the $n+1$ th patch should be left when the reward rate falls to $\gamma_{n}$. Thus in the terminology of Appendix 1

$$
R_{n+1}=R\left(t\left(\gamma_{n}\right)\right)
$$

and

$$
T_{n+1}=\tau+t\left(\gamma_{n}\right)
$$

Thus by equation (A1.5)

$$
\begin{equation*}
R_{n+1}=f\left(\gamma_{n}\right) T_{n+1} \tag{A2.1}
\end{equation*}
$$

For convenience we set

$$
X_{n}=\sum_{i=0}^{n} T_{i} \quad \text { and } \quad Y_{n}=\sum_{i=0}^{n} R_{i}
$$

so that

$$
\begin{equation*}
\gamma_{n}=\frac{Y_{n}}{X_{n}} \tag{A2.2}
\end{equation*}
$$

We also have

$$
\gamma_{n+1}=\frac{Y_{n}+R_{n+1}}{X_{n}+T_{n+1}}
$$

and hence

$$
\begin{equation*}
\gamma_{n+1}=\frac{\gamma_{n} X_{n}+f\left(\gamma_{n}\right) T_{n+1}}{X_{n}+T_{n+1}} \tag{A2.3}
\end{equation*}
$$

by equations (A2.1) and (A2.2).

Suppose that $\gamma_{n}>\gamma^{*}$. Then $f\left(\gamma_{n}\right) \leq \gamma^{*}\left(<\gamma_{n}\right)$ by equation (2), and hence $\gamma_{n} X_{n}+f\left(\gamma_{n}\right) T_{n+1}<\gamma_{n}\left(X_{n}+T_{n+1}\right)$. Equations (A2.3) then shows that $\gamma_{n+1}<\gamma_{n}$. We also have $\gamma_{n+1}>0$ since $X_{n}$ and $\gamma_{n}$ are positive. This proves equation (6).

Suppose $\gamma_{n}=\gamma^{*}$. The $f\left(\gamma_{n}\right)=\gamma^{*}$ by equation (3) and hence $\gamma_{n+1}=\gamma^{*}$ by equation (A2.3). This proves equation (7).

Finally, suppose $0<\gamma_{n}<\gamma^{*}$. Then $f\left(\gamma_{n}\right)>\gamma_{n}$ by equation (4), and $\gamma^{*} \geq$ $f\left(\gamma_{n}\right)$ by equation (2). Thus

$$
\gamma_{n}\left(X_{n}+T_{n+1}\right)<\gamma_{n} X_{n}+f\left(\gamma_{n}\right) T_{n+1}<\gamma^{*}\left(X_{n}+T_{n+1}\right)
$$

Equation (8) then follows from this equation and equation (A2.3).

## APPENDIX 3

## Mean Square Error

The function $f(\gamma)$ has a maximum at $\gamma=\gamma^{*}$. Thus

$$
\begin{equation*}
f^{\prime}\left(\gamma^{*}\right)=0 \tag{A3.1}
\end{equation*}
$$

and

$$
\begin{equation*}
k \equiv-\frac{1}{2} f^{\prime \prime}\left(\gamma^{*}\right) \geq 0 \tag{A3.2}
\end{equation*}
$$

Performing a Taylor series expansion of $f(\gamma)$ about $\gamma=\gamma^{*}$ we obtain

$$
f(\gamma)=f\left(\gamma^{*}\right)+\left(\gamma-\gamma^{*}\right) f^{\prime}\left(\gamma^{*}\right)+\frac{1}{2}\left(\gamma-\gamma^{*}\right)^{2} f^{\prime \prime}\left(\gamma^{*}\right)+0\left(\left(\gamma-\gamma^{*}\right)^{3}\right)
$$

Thus by equations (A3.1) and (A3.2)

$$
f\left(\gamma^{*}\right)-f(\gamma)=k\left(\gamma-\gamma^{*}\right)^{2}+0\left(\left(\gamma-\gamma^{*}\right)^{3}\right)
$$

This equation shows that the loss in long term reward rate which results from using the estimate $\gamma$ rather than $\gamma^{*}$ is proportional to $\left(\gamma-\gamma^{*}\right)^{2}$ for small $\left|\gamma-\gamma^{*}\right|$.

