J. theor. Biol. (2003) **220,** 169–188
doi:10.1006/jtbi.2003.3146, available online at http://www.idealibrary.com on IDEAL®

Carving the Cognitive Niche: Optimal Learning Strategies in Homogeneous and Heterogeneous Environments

Benjamin Kerr*† and Marcus W. Feldman†

†*Department of Biological Sciences*, *Stanford University*, *Stanford*, *CA* 94305, *U.S.A.*

(*Received on* 7 *September* 2001, *Accepted in revised form on* 12 *August* 2002)

A model learning system is constructed, in which an organism samples behaviors from a behavioral repertoire in response to a stimulus and selects the behavior with the highest payoff. The stimulus and most rewarding behavior may be kept in the organism's long-term memory and reused if the stimulus is encountered again. The value of the memory depends on the reliability of the stimulus, that is, how the corresponding payoffs of behaviors change over time. We describe how the inclusion of memory can increase the optimal sampling size in environments with some stimulus reliability. In addition to using memory to guide behavior, our organism may use information in its memory to choose the stimulus to which it reacts. This choice is influenced by both the organism's memory state and how many stimuli the organism can observe (its sensory capability). The number of sampled behaviors, memory length, and sensory capability are the variables that define the learning strategy. When all stimuli have the same reliability, there appears to be only a single optimal learning strategy. However, when there is heterogeneity in stimulus reliability, multiple locally optimal strategies may exist.

1. Introduction

Several verbal and mathematical treatments of the evolutionary advantages of learning suggest that the amount of environmental variability is key (Arnold, 1978; Plotkin & Odling-smee, 1979; Johnston, 1982; Stephens, 1991; Bergman & Feldman, 1995; Feldman *et al.*, 1996; Dukas, 1998a; Ancel, 1999). If there is too little variability, then one would expect a "hard-wired" behavior appropriate for the constant environment. If there is too much variability, then learned behaviors rapidly lose their utility and learning is a costly waste of time. It seems that the variability must be "just right;" a sort of "Goldilocks principle" applies. Dukas (1998a) sums up this principle when he writes that while learning is advantageous when there is some pattern of environmental variation, "the learning rate (must) be sufficiently higher than the rate of environmental change."

What exactly is meant by environmental variability? Here we frame the question in terms of a stimulus and a response. If the same response to a given stimulus has different values at two points in time, one might argue that a change in the environment has occurred between these points in time. Specifically, an organism might visit the same object twice, with some transformation (unperceived by the organism) in the object occurring between visits that affects

*Corresponding author. Present address: Department of Ecology, Evolution and Behavior, University of Minnesota, 100 Ecology, 1987 Buford Circle, St. Paul, MN 55108, U.S.A. Tel.: +1-612-624-3769.
E-mail address: bkerr@leland.stanford.edu (B. Kerr).

0022-5193/03/$35.00

the value of a given response. Alternatively, the organism might visit two distinct objects, both perceived as the same stimulus. If each object yields a different value for a given response, the "single" stimulus appears to change. If the value of the same response drops on average when the "environment changes," the value of learning will be adversely affected by high rates of change. One might expect that the value of the stimulus/response pair needs to remain *constant enough* in order for learning to be favored. This simply restates one component of the Goldilocks principle.

A complication arises if the organism lives in an environment with many stimuli, each possessing its own dynamics. These different stimuli may have different rates of change. That is, there may be *heterogeneity* in environmental variability. Heterogeneity plays heavily into Godfrey-Smith's (1998) environmental complexity thesis, which states: "The function of cognition is to enable the agent to deal with environmental complexity." Godfrey-Smith makes it clear that he has heterogeneity in mind when he speaks of complexity. Stephens (1987,1991) has modeled systems where resources are heterogeneous in their variability, while other authors (including Krebs *et al.*, 1978; Cohen, 1991) have explored models in which foraging organisms must obtain information about heterogeneous patch quality through sampling. How does such heterogeneity influence the value of certain forms of learning? What general tenet, if any, will replace the Goldilocks principle? In this paper, we attempt to build a framework that extends these earlier modeling efforts (Krebs *et al.*, 1978; Stephens 1987, 1991; Cohen, 1991) whereby we can rigorously explore these questions.

2. Optimality: Analytic Model and Results

2.1. SAMPLING

Our model organism's life is modeled as a series of discrete time steps. In this section we consider a single time step in which an organism reacts to an unfamiliar stimulus. Suppose that there are $n \geqslant 2$ behaviors available to an organism in response to this stimulus (Harley & Maynard Smith, 1983); we denote the set of behaviors as $\mathrm{B} = \{\beta_1, \beta_2, \beta_3, \ldots, \beta_n\}$. Depending on the stimulus, these behaviors might represent different ways to obtain food, mates, nesting territory, etc. Each behavior carries a different numerical payoff value. For instance, if the organism is foraging, the value might represent caloric intake. We assume that payoff value influences the fitness of the organism; a higher payoff value corresponds to higher fitness.

When the stimulus is unfamiliar, the time step is broken into two periods: a sampling period and a selection period. During the sampling period, the organism tests a sequence of z behaviors ($z \geqslant 1$) and retains their payoff values in working memory. Here we assume that each behavior in the sequence is independently sampled with replacement from the n equally likely possibilities. Consequently, z could be less than, equal to, or greater than n. During the selection period, the organism chooses the behavior with the highest associated value and receives the corresponding payoff.

We assume that the sampling period supplies the organism with *information* on the sampled behaviors and that the payoff occurs during the selection period. Thus, while sampling more behaviors will, on average, lead to a better selection, a longer sampling period leads to a shorter selection period and thus less time to reap the benefit. In other words, the payoff over the time step is not only a function of the behavior chosen, but also a function of the length of the sampling and selection periods. We introduce this feature into the model via a sampling cost. Specifically, the fraction of the time step taken up by sampling is cz, where $c > 0$ is interpreted as the fraction of the time step that any sampled behavior consumes. (Note that $1 \leqslant z < 1/c$, such that $1/c$ is the maximum number of behaviors an organism can sample in a single time step.)

So, time spent sampling is time taken away from using the best behavior to reap the corresponding payoff. Lewis (1986) showed that cabbage butterflies that were not familiar with a certain flower species took a certain amount of time to discover nectar. Once the location of nectar was learned (i.e., the appropriate behavior to access nectar was discovered) this "discovery time" dropped dramatically with future trips to the same species. Thus, there may be a time investment in sampling behaviors in response to an unfamiliar stimulus.

In our model, we assume each of the n behaviors in the set B is uniquely assigned a payoff value from the set $\Pi = \{\pi_1, \pi_2, \pi_3, \ldots, \pi_n\}$, where $\pi_i < \pi_j$ if $i < j$. That is, there is a one-to-one function $\Phi : \mathrm{B} \mapsto \Pi$, where $\Phi(\beta) = \pi$ with $\beta \in \mathrm{B}$ and $\pi \in \Pi$. We assume that the organism does not know the range of Π or the value of n; thus, despite the payoff of a current sampled behavior, there is always the chance that the organism may sample a behavior with a higher payoff. Let the payoff that is associated with the most rewarding behavior of the z sampled behaviors be given by π_{max}. The probability that $\pi_{max} = \pi_k$, $P_{\pi_{max}=\pi_k}(z)$, is given by

$$P_{\pi_{max}=\pi_k}(z) = P_{\pi_{max}\leqslant\pi_k}(z) - P_{\pi_{max}\leqslant\pi_{k-1}}(z).$$

Because all z sampled behaviors are i.i.d. uniform random variables,

$$P_{\pi_{max}\leqslant\pi_k}(z) = \left(\frac{k}{n}\right)^z.$$

Therefore, the probability that a sampling individual chooses a behavior with payoff π_k (we abbreviate $P_{\pi_{max}=\pi_k}(z)$ as $P_{\pi_k|\mathbf{S}}(z)$, where the subscript **S** indicates that sampling occurred) is

$$P_{\pi_k|\mathbf{S}}(z) = \frac{k^z - (k-1)^z}{n^z}, \qquad (1)$$

and the net payoff value obtained by an organism with sampling size z is given by

$$V_{\pi_k|\mathbf{S}}(z) = (1 - cz)\pi_k. \qquad (2)$$

Note that if the organism samples z behaviors, then it has the fraction $(1-cz)$ of the time step to reap the benefit of the chosen behavior. Using eqns (1) and (2), the expected payoff value for sampling z behaviors is

$$\begin{aligned}\bar{V}_{\mathbf{S}}(z) &= \sum_{k=1}^{n}(V_{\pi_k|\mathbf{S}}(z))(P_{\pi_k|\mathbf{S}}(z)) \\ &= \frac{(1-cz)}{n^z}\sum_{k=1}^{n}[k^z - (k-1)^z]\pi_k. \qquad (3)\end{aligned}$$

For simplicity, we will make the assumption that $\pi_k = (k-1)/(n-1)$, such that the payoffs range between 0 and 1 per time step. Given this assumption, we have

$$\bar{V}_{\mathbf{S}}(z) = (1 - cz)\left(1 - \frac{1}{n-1}\sum_{k=1}^{n-1}\left(\frac{k}{n}\right)^z\right). \qquad (4)$$

With z arbitrary, no general closed form for $\sum_{k=1}^{n-1}(k/n)^z$ exists, although a recursive relationship can be defined (Beardon, 1996).

2.2. OPTIMAL SAMPLING

To find the sampling size that produces the maximum expected payoff value, we treat expression (4) as a continuous function in z and consider $(\partial\bar{V}_{\mathbf{S}}(z)/\partial z) = 0$, which gives

$$c = \frac{\sum_{k=1}^{n-1}\ln(n/k)(k/n)^z}{n - 1 + \sum_{k=1}^{n-1}\{\ln(n/k)^z - 1\}(k/n)^z}. \qquad (5)$$

Since the right-hand side of eqn (5) approaches zero as $z \to \infty$ and is monotone decreasing in z, there is a single critical point for $\bar{V}_{\mathbf{S}}(z)$, a global maximum.

For a given value of c, we denote the z value that solves expression (5) as z^*_{opt}, which is a positive real number greater than unity if the right-hand side of eqn (5) exceeds c at $z = 1$. From eqn (5), it can be shown that z^*_{opt} monotonically decreases with c (i.e., optimal sampling size decreases with increasing sampling cost). Since there is only a single critical point (the maximum), either the integer immediately above or below z^*_{opt} is the relevant maximum for $\bar{V}_{\mathbf{S}}(z)$ if only integer values are allowed for z.

2.3. SAMPLING WITH MEMORY (LEARNING)

Suppose our organism lives for $T \geqslant 2$ time steps, and denote the current time step by t $(1 \leqslant t \leqslant T)$. Each time step could conceivably be filled with a bout of sampling and then the selection of a behavior to use. However, the lifetime of an organism is not necessarily the concatenation of unrelated bouts of sampling. Rather, if the most valuable behavior of the sample is stored in long-term memory, this behavior might be reused in an appropriate situation. By relying on past experience, *the organism may eschew the sampling process and its associated costs*. In Lewis' (1986) study of cabbage butterflies, the time to discover nectar decreases in successive visits to the same flower species because, presumably, the butterfly has recorded information (location of nectar or a method of nectar extraction) into its memory and reuses it appropriately. Thus, memory may relieve the cost of sampling.

We envision the complete environment of an organism as a collection of $N \geqslant 2$ stimuli, where the set of stimuli is given by $\sum = \{\sigma_1, \sigma_2, \sigma_3, \ldots, \sigma_N\}$. For each stimulus, assume there are n behaviors available to the organism. We will assume that for each stimulus $\sigma \in \sum$, there exists a one-to-one function $\Phi_\sigma : \mathrm{B} \mapsto \Pi$, associating a unique payoff value with each behavior. In this section, we assume that at each time step, the organism encounters one of the N stimuli at random. When the stimulus is novel, the organism samples z behaviors ($z \geqslant 1$) and picks the most valuable behavior. Now, when this organism returns to a stimulus that it has experienced in the past, it may do one of the following: (i) sample z behaviors and pick the most valuable or (ii) reuse the behavior chosen from a previous sampling bout with that stimulus. We imagine that option (i) will occur if the organism sampled in its previous encounter but simply forgets the behavior it chose. Option (ii) will occur when the organism sampled in a previous encounter with the stimulus and remembers the behavior it chose; that is, the organism is able to retrieve the memory. (Note that in Appendix B we consider organisms that will use option (i) for a *familiar* stimulus when the remembered payoff for the behavior in memory is too low—i.e., occasional resampling). If sampling is costly, option (ii) would offer a benefit—a behavior is initiated without the costly sampling process.

In this paper, we take an extremely simplified view of long-term memory. For each time step, the behavior employed by the organism is recorded into memory along with the stimulus to which the organism reacted. This memory has a maximum lifetime of m time steps (where $0 \leqslant m \leqslant T$). The variable m is considered the long-term memory length. One can imagine that when an organism reacts to a stimulus, a "timer" is set that will go off after m time steps. If the stimulus is not seen again within m time steps (i.e., the timer goes off), then the stimulus–behavior pair in memory is lost. If the stimulus is revisited within m time steps, the memory is renewed (i.e., the timer is reset). As a consequence of this memory structure, the organism can remember at most m stimulus–behavior pairs at any one time step. An explicit cost to long-term memory, c_m, might be imagined (e.g., some metabolic cost); however, in this paper, we will assume $c_m = 0$.

Although memory allows an organism to reuse behaviors in response to a familiar stimulus and avoid a costly sampling process, the success of remembered behaviors depends on the constancy of behavioral payoff values. That is, the probability that the payoff of the used behavior has changed between the time step at which the behavior was originally sampled and the time step of its reuse will influence the value of memory. Here, we imagine that the payoffs for the behaviors in the response repertoire to stimulus $\sigma \in \sum$ change with probability ρ_σ at every time step. Thus, the parameter ρ_σ relates to the stimulus reliability, with $\rho_\sigma = 0$ corresponding to complete reliability (the payoffs of behaviors never change) and $\rho_\sigma = 1$ meaning complete unreliability (the payoffs of behaviors changing every time step).

Change in stimulus σ is accommodated by allowing the mapping from behaviors to payoffs, $\Phi_\sigma : \mathrm{B} \mapsto \Pi$, to change over time steps (thus, Φ_σ becomes time-dependent). We model stimulus change as follows: Consider some permutation function on the set of behaviors, $\Psi_j : \mathrm{B} \mapsto \mathrm{B}$. There are $n!$ Ψ functions, which we arbitrarily label $\Psi_1, \Psi_2, \Psi_3, \ldots, \Psi_{n!}$. Then we have the following:

$$\Phi_\sigma(t+1) = \begin{cases} \Phi_\sigma(t) & \text{if there is no change to stimulus } \sigma \text{ at } t, \\ \Phi_\sigma(t) \circ \Psi_J & \text{if there is a change to stimulus } \sigma \text{ at } t \end{cases} \tag{6}$$

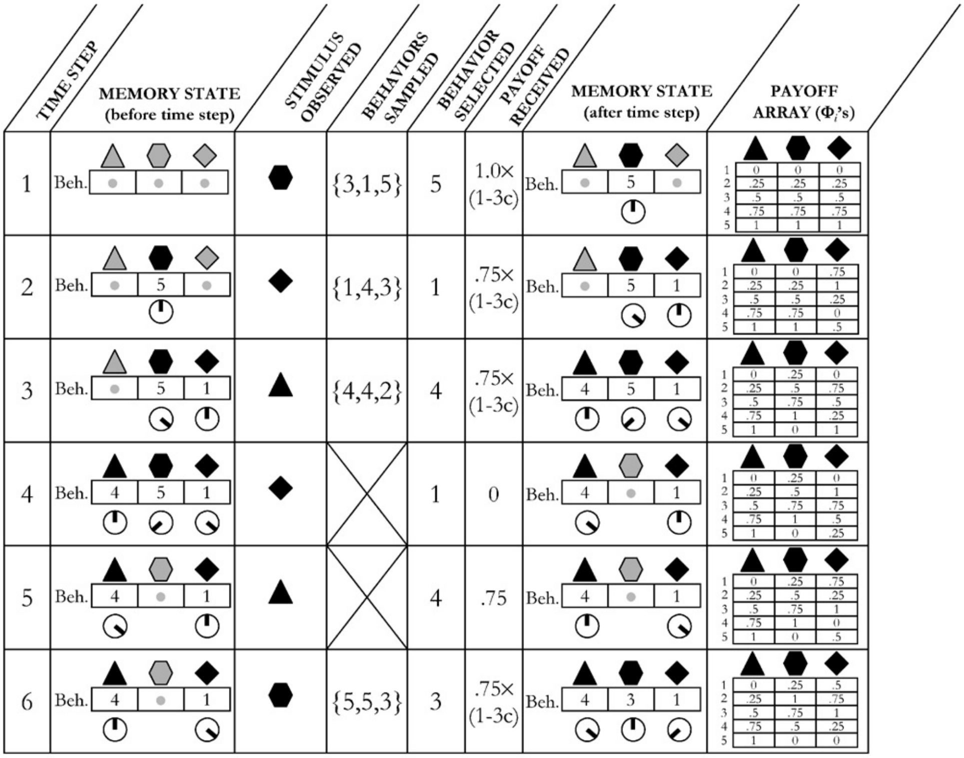


FIG. 1. Learning by an organism with $z = 3$ and $m = 3$. The three stimuli ($N = 3$) in the environment are triangle, hexagon, and diamond. There are five behaviors ($n = 5$) the organism can use in response to each stimulus. Every time step, the organism observes one of the stimuli (chosen at random). If the stimulus is not in memory, the organism will sample three behaviors and record a stimulus–behavior pair into its memory (time steps 1, 2, 3 and 6 in the figure). It will receive the payoff of the most valuable behavior (multiplied by a factor for the sampling cost) and will record this behavior into its memory. In the columns marked "MEMORY STATE," the cognitive mapping is given as a single row table with the stimulus over the behavior recorded in response to it (a gray circle appears under the stimulus when the organism carries no response to that stimulus in memory). In the cognitive mapping, the stimulus is black when the organism has it in memory and is gray otherwise. The memory length of the organism can be visualized by a "timer" being set each time the organism reacts to a given stimulus (the small clock underneath the stimulus–behavior pair). This timer expires after three time steps (in the figure, it ticks clockwise starting from 12:00 to 4:00, then 4:00 to 8:00 and finally 8:00 back 12:00) and if the stimulus is not revisited within three time steps, the stimulus–behavior pair is lost. For instance, the hexagon stimulus is experienced at time step 1, but is lost at time step 4 since it was not experienced from time steps 2 to 4. If the experienced stimulus exists in memory, the organism will reuse the behavior recorded in memory (time steps 4 and 5) and then reset its timer for that stimulus/behavior pair. Note that the sampling cost ($3c$) is avoided when memory is used. The last column in the figure gives the array of payoff values (the Φ mappings). Specifically, the matrix is filled with the payoff of each behavior (the rows) in response to each stimulus (the columns). In the figure, the reliabilities of the stimuli differ. The triangle has $\rho = 0$ (that is, the payoff of each behavior never changes), the hexagon has $\rho = 0.5$ (the payoffs undergo a random permutation on the behaviors about 50% of the time), and the diamond has $\rho = 1$ (the payoffs permute every time step). As a consequence of these reliabilities, one can see that the reliable stimulus (the triangle) always returns a consistent payoff when the same behavior is remembered (compare time step 3 with time step 5) whereas the unreliable stimulus (the diamond) returns an inconsistent payoff for remembered behaviors (compare time step 2 with time step 4).

with J a discrete random variable with $J \sim Unif(1, n!)$ and where "$\circ$" represents functional composition. Figure 1 illustrates stimulus reliability and the memory structure with an example of an individual behaving over several time steps.

2.4. OPTIMAL LEARNING

2.4.1. *Derivation of Expected Lifetime Payoff*

We represent the sequence of payoffs over the organism's lifetime by the quantity $\pi_{\vec{\kappa}}$, with $\vec{\kappa} = \langle k_1, k_2, k_3, ..., k_t, ..., k_T \rangle$, such that the payoff of a behavior employed at time step t is given by $\pi_{k_t} = (k_t - 1)/(n - 1)$. The lifetime payoff value, $LV_{\pi_{\vec{\kappa}}}(z, m)$, is the sum of the payoffs across time steps. Specifically, if $V_{\pi_{k_t}}(z, m)$ gives the net payoff at time step t to an organism with sampling size z and memory length m, we have

$$LV_{\pi_{\vec{\kappa}}}(z, m) = \sum_{t=1}^{T} V_{\pi_{k_t}}(z, m).$$

In this section, we deduce an expression for the expected lifetime payoff value of an organism, $\overline{LV}(z, m) = \sum_{t=1}^{T} \sum_{k_t=1}^{n} V_{\pi_{k_t}}(z, m)(P_{\pi_{k_t}}(z, m))$, where $P_{\pi_{k_t}}(z, m)$ is the probability of employing a behavior with a payoff of π_{k_t} at time step t (either from memory or from a sampling bout). To this end, it helps to consider the possible events that can occur at any time step for a given organism. Let **S** be the event that an organism samples in response to a stimulus, **M** be the event that an organism remembers a stimulus, **C** be the event that an organism reuses a consistent behavior from memory and **I** be the event that an organism reuses an inconsistent behavior from memory. By "consistent" ("inconsistent") we mean that the payoffs of the behaviors have not been (have been) randomly permutated since the organism originally recorded the stimulus and behavior into memory and therefore, the remembered behavior has (may not have) the same payoff as when it was originally stored into memory. Note that one of the mutually exclusive events **S**, **C**, or **I** must occur each time step (if **C** or **I** occurs, then **M** must occur). Below we let **A** denote one of events **S**, **C**, or **I**.

We can now give the general expression for the expected lifetime payoff:

$$\overline{LV}(z, m) = \sum_{t=1}^{T} \sum_{k_t=1}^{n} \sum_{\mathbf{A} \in \{\mathbf{S},\mathbf{C},\mathbf{I}\}} (V_{\pi_{k_t}|\mathbf{A}}(z, m)) \times (P_{\pi_{k_t}|\mathbf{A}}(z, m)) P_{\mathbf{A}}(t, z, m).$$

If we know that event **S**, **C**, or **I** occurred at time step t then the probability and net payoff value of employing a behavior that gives a payoff of $\pi_{k_t} = (k_t - 1)/(n - 1)$ should not depend on memory size or the time step. Also, the probability of a certain event does not depend on the sampling size. So we can rewrite the above as

$$\begin{aligned}\overline{LV}(z, m) &= \sum_{t=1}^{T} \sum_{\mathbf{A} \in \{\mathbf{S},\mathbf{C},\mathbf{I}\}} \left[\left\{ \sum_{k=1}^{n} (V_{\pi_k|\mathbf{A}}(z))(P_{\pi_k|\mathbf{A}}(z)) \right\} \times P_{\mathbf{A}}(t, m) \right] \\ &= \sum_{t=1}^{T} \sum_{\mathbf{A}=\{\mathbf{S},\mathbf{C},\mathbf{I}\}} (\bar{V}_{\mathbf{A}}(z))(P_{\mathbf{A}}(t, m)), \qquad (7)\end{aligned}$$

where $\bar{V}_{\mathbf{A}}(z)$ gives the expected payoff given event **A**, and $P_{\mathbf{A}}(t, m)$ gives the probability of event **A**.

Equation (4) gives $\bar{V}_{\mathbf{S}}(z)$. If the organism reuses a consistent behavior from memory, its expected payoff is the same as if it had sampled without any cost, namely

$$\bar{V}_{\mathbf{C}}(z) = \frac{\bar{V}_{\mathbf{S}}(z)}{1 - cz}. \qquad (8)$$

If the organism reuses an inconsistent behavior from memory, there has been a random permutation on the behavior to payoff map [see eqn (6)]. Therefore, the expected payoff of an inconsistent behavior is identical to that given by sampling a single randomly chosen behavior without the cost, namely

$$\bar{V}_{\mathbf{I}}(z) = \frac{\bar{V}_{\mathbf{S}}(1)}{1 - c}. \qquad (9)$$

In order to compute the probabilities of **S**, **C**, and **I**, we must first consider the probability that the organism uses memory (event **M**). This probability is

$$P_{\mathbf{M}}(t, m) = 1 - \left(\frac{N - 1}{N}\right)^{\min(m, t-1)}. \qquad (10)$$

Note that $P_{\mathbf{M}}(t,0) = 0$ and $P_{\mathbf{M}}(1,m) = 0$, for all t and m (i.e., without memory the organism must sample and even with memory, the organism must sample for its first time step). To verify (10), consider an organism at time step t. The last d time steps over which the organism can remember is $d = \min(m, t - 1)$. One of the N stimuli is experienced (at random) at time step t. The probability that this stimulus has not been experienced in the past d time steps is simply

$[(N-1)/N]^d$ since the stimuli are experienced independently and with equal probability every time step. Thus, the probability that the organism remembers the stimulus is $1-[(N-1)/N]^d$.

Since event **S** is the complement of the event **M**, we have

$$P_{\mathbf{S}}(t,\, m) = 1 - P_{\mathbf{M}}(t,\, m) = \left(\frac{N-1}{N}\right)^{\min(m,t-1)}. \quad (11)$$

We use the law of total probability to deduce formulas for $P_{\mathbf{C}}(t,m)$ and $P_{\mathbf{I}}(t,\, m)$:

$$P_{\mathbf{C}}(t,m) = (P_{\mathbf{C}|\mathbf{M}}(t,\, m))P_{\mathbf{M}}(t,\, m), \quad (12)$$

$$\begin{aligned} P_{\mathbf{I}}(t,\, m) &= (P_{\mathbf{I}|\mathbf{M}}(t,\, m))P_{\mathbf{M}}(t,\, m) \\ &= (1 - P_{\mathbf{C}|\mathbf{M}}(t,\, m))P_{\mathbf{M}}(t,\, m), \quad (13) \end{aligned}$$

where $P_{\mathbf{A}|\mathbf{B}}(t,\, m)$ is the conditional probability of event **A** given **B** at time step t for an organism with memory size m. To derive eqns (12) and (13), note that if the organism does not remember a stimulus, then it cannot use a consistent or inconsistent behavior from memory. Consequently, $P_{\mathbf{C}|\mathbf{S}}(t,\, m) = P_{\mathbf{I}|\mathbf{S}}(t,\, m) = 0$, where event **S** is the complement of event **M**, and eqns (12) and (13) result. The conditional probability of event **C** (or **I**) given event **M** is not defined when $t=1$ or $m=0$ (since event **M** does not occur in such cases). We have $P_{\mathbf{C}}(1,\, m) = P_{\mathbf{I}}(1,\, m) = 0$ for all m and $P_{\mathbf{C}}(t,\, 0) = P_{\mathbf{I}}(t,\, 0) = 0$ for all t, since $\mathbf{M} = \mathbf{C} \cup \mathbf{I}$ and $P_{\mathbf{M}}(1,\, m) = P_{\mathbf{M}}(t,\, 0) = 0$. When $P_{\mathbf{M}}(t,m) \neq 0$, we have $P_{\mathbf{I}|\mathbf{M}}(t,\, m) = 1 - P_{\mathbf{C}|\mathbf{M}}(t,\, m)$ [see eqn (13)], since $\mathbf{C} \cap \mathbf{I} = \varnothing$.

If we assume that all stimuli have the same reliability (i.e., $\rho_\sigma = \rho$ for all $\sigma \in \sum$), we have

$$\begin{aligned} P_{\mathbf{C}|\mathbf{M}}(t,\, m) = &\sum_{i=1}^{\min(m,t-1)} \{\Lambda(i,\, t,\, m) \\ &\times [P_{\mathbf{C}}(t-i,\, m) + P_{\mathbf{S}}(t-i,\, m)](1-\rho)^i\}. \quad (14) \end{aligned}$$

with $\Lambda(i,\, t,\, m) = ((N-1)^{i-1} N^{\min(m,t-1)-i}) / \sum_{i=1}^{\min(m,t-1)} (N-1)^{i-1} N^{\min(m,t-1)-i}$.

We derive eqn (14) as follows. At time step t, since we condition on event **M**, the current stimulus must have been experienced in one of the $d = \min(m,\, t-1)$ past time steps. Suppose that the stimulus currently experienced was most recently experienced i time steps ago (at time step $t-i$ with $1 \leqslant i \leqslant d$). The probability that the stimulus has not changed since the most recent encounter is $(1-\rho)^i$. The probability that the stimulus was unfamiliar when recorded in memory at time step $t-i$ is given by $P_{\mathbf{S}}(t-i,\, m)$. However, we must also consider the possibility that the organism used its memory at time step t–i. The probability that the organism used memory at time step $t-i$ and employed *a consistent behavior* is $P_{\mathbf{C}}(t-i,\, m)$. Thus, the probability that an organism remembers a consistent behavior, given that it remembers a stimulus/behavior pair from i time steps ago, is $[P_{\mathbf{S}}(t-i,\, m) + P_{\mathbf{C}}(t-i,\, m)](1-\rho)^i$. To compute $P_{\mathbf{C}|\mathbf{M}}(t,\, m)$, we must sum $[P_{\mathbf{S}}(t-i,\, m) + P_{\mathbf{C}}(t-i,\, m)](1-\rho)^i$ across the d time steps. However, not all of the d time steps are equally likely. The proportion of sequences in which the current stimulus is experienced most recently i time steps ago is given by $\Lambda(i,\, t,\, m)$. Thus, the probability $P_{\mathbf{C}|\mathbf{M}}(t,\, m)$ is derived by summing over i and weighting each of the d time steps accordingly.

Given the parameters ρ and N, as well as the memory size m, we can recursively compute $P_{\mathbf{S}}(t,\, m)$, $P_{\mathbf{C}}(t,\, m)$, or $P_{\mathbf{I}}(t,\, m)$ for any value of t given that $P_{\mathbf{S}}(1,\, m) = 1$, $P_{\mathbf{M}}(1,\, m) = 0$, $P_{\mathbf{C}}(1,\, m) = 0$ and $P_{\mathbf{I}}(1,\, m) = 0$. These probabilities are used along with the expectation of the conditional payoffs [eqns (4), (8) and (9)] to compute the expected lifetime payoff for a given $(z,\, m)$ combination. We then can compute the optimal sampling size and memory length for specific environmental conditions (c and ρ) by searching the expected lifetime payoff surface for maxima.

2.4.2. *Optimal Sampling Revisited*

For a given memory and stimulus reliability, we find the sampling size that maximizes expected lifetime payoff in eqn (7) by setting $(\partial \overline{LV}(z,\, m)/\partial z) = 0$, which, after rearranging and simplifying, is equivalent to

$$c = (\alpha)\frac{\sum_{k=1}^{n-1} \ln(n/k)(k/n)^z}{n-1+\sum_{k=1}^{n-1}\{\ln(n/k)^z - 1\}(k/n)^z}, \quad (15)$$

with $\alpha = \sum_{t=1}^{T}[P_{\mathbf{C}}(t,\, m) + P_{\mathbf{S}}(t,\, m)] / \sum_{t=1}^{T} P_{\mathbf{S}}(t,\, m)$.

Equation (15) is identical to eqn (5) except for the α factor. If $P_C(t, m) = 0$ for all t, which will happen if and only if $m = 0$ or $\rho = 1$ [see eqns (10), (12) and (14)], then $\alpha = 1$. In such a case, the z^*_{opt} from eqn (5) will be identical to the z that solves eqn (15), which we will call z_{opt}. So, without memory or if the stimuli change every time step, the organism should use the sampling size that maximizes the expected payoff of a single sampling bout (i.e., $z_{opt} = z^*_{opt}$). On the other hand, if the organism has some memory ($m > 0$) and the stimuli are not completely unreliable ($0 \leqslant \rho < 1$), then $P_C(t, m) > 0$ for some t, and $\alpha > 1$. If $\alpha > 1$, then $z_{opt} > z^*_{opt}$, since the right-hand side of eqn (5) is monotone decreasing in z. Thus, with some memory and some stimulus reliability, it may pay to invest in a higher sampling size than that which maximizes payoff over a single time step (see Fig. 2). The potential difference between z_{opt} and z^*_{opt} illustrates an important point: sampling for long-term gain is a different exercise than sampling for "myopic" gain (*sensu* Mangel & Clark, 1988). This echoes a result in dynamic programming models, where optimal decisions made early in life (potentially concerning the long-term condition of the organism) differ from optimal decisions made late in life (when the organism may be more short-sighted).

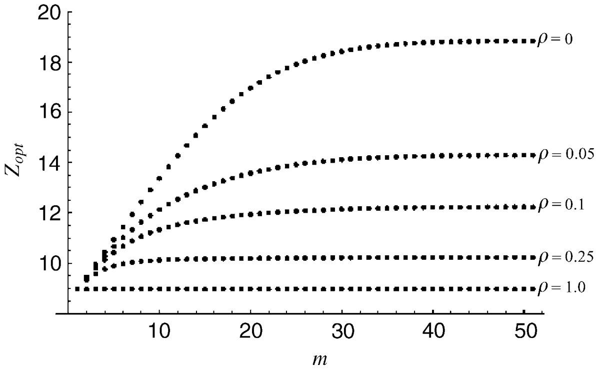


FIG. 2. The z values that solve eqn (15) are plotted against a fixed memory value for the case of $c = 0.01$, $n = 10$, $T = 50$, and $N = 10$. Note that if $\rho < 1$, the optimal sampling size increases with memory. As ρ decreases (the stimuli become more reliable), the optimal sampling size increases. When $m = 0$ or $\rho = 1$, the z value that solves eqn (15) is identical to the z value that solves eqn (5) (i.e., $z_{opt} = z^*_{opt}$). Otherwise, $z_{opt} > z^*_{opt}$.

2.4.3. *Optimal Memory Length*

If the sampling size, z, is fixed, numerical methods using eqns (7)-(14) give the memory length, m_{opt}, that maximizes eqn (7). This optimal memory is shown as a function of z in Fig. 3. As the stimuli become less reliable (i.e., ρ increases), optimal memory decreases for any given sampling size, as is expected. As sampling size (z) increases, optimal memory eventually increases. With regard to memory use, there are two essential considerations. First, what is the cost of using an inconsistent behavior vs. the benefit of using a consistent behavior? Second, what is the probability that a reused behavior from memory is consistent?

Figure 4 illustrates both considerations with a graph of the expected payoff of a sampling bout with sampling size z, $\bar{V}_S(z)$, plotted with the expected payoff value of a consistent memory, $\bar{V}_C(z)$, and an inconsistent memory, $\bar{V}_I(z)$. When compared with sampling, the potential benefit of memory (using a consistent behavior) is given by the difference between $\bar{V}_C(z)$ and $\bar{V}_S(z)$, while the potential cost of memory (using an inconsistent behavior) is given by the difference between $\bar{V}_S(z)$ and $\bar{V}_I(z)$. We see that the benefit grows with z, while the cost at first grows with z and then decreases with z. If $z > \check{z}$ in Fig. 4, the benefit of memory grows and the cost of memory decreases as z increases. Thus, optimal memory size should either stay constant or increase as z increases from $\check{z}$. As ρ decreases, the stimuli become more reliable and the organism is more

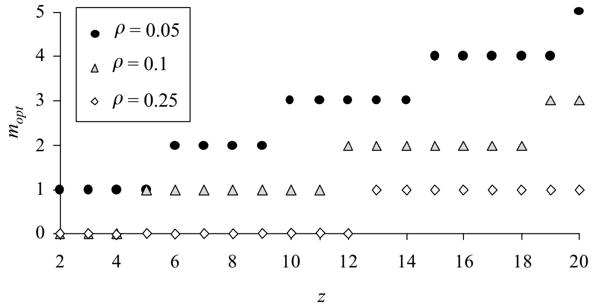


FIG. 3. The m values that optimize eqn (7) are plotted against a fixed sampling size for the case of $c = 0.01$, $n = 10$, $T = 50$, and $N = 10$. In general, the optimal memory will increase with increased sampling size. Also, as ρ decreases (the stimuli becoming more reliable), the optimal memory will increase.

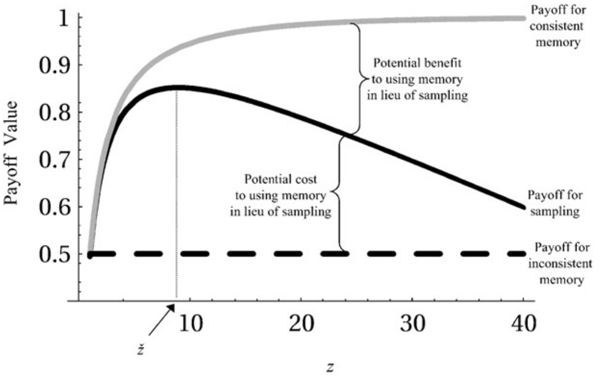


FIG. 4. Three functions of sampling size are shown. The first is a horizontal dashed line [$\bar{V}_{\mathbf{I}}(z)$] and gives the expected payoff for an inconsistent behavior from memory. The second is the solid black function and gives the expected payoff, including cost ($c = 0.01$), of sampling z behaviors [$\bar{V}_{\mathbf{S}}(z)$]. The third is the solid gray function and gives the expected payoff for a consistent behavior from memory [$\bar{V}_{\mathbf{C}}(z)$]. Thus, the difference between the gray and black functions gives the potential benefit of using memory over sampling, while the difference between the black and dashed functions gives the potential cost of using memory. As z increases past the value marked z' memory size should not decrease (since the benefit grows and the cost shrinks). Eventually $\bar{V}_{\mathbf{S}}(z)$ will cross $\bar{V}_{\mathbf{I}}(z)$. If $\check{z}$ is this crossing value (note $\acute{z} < 1/c$), we will consider only $1 \leqslant z < \acute{z}$.

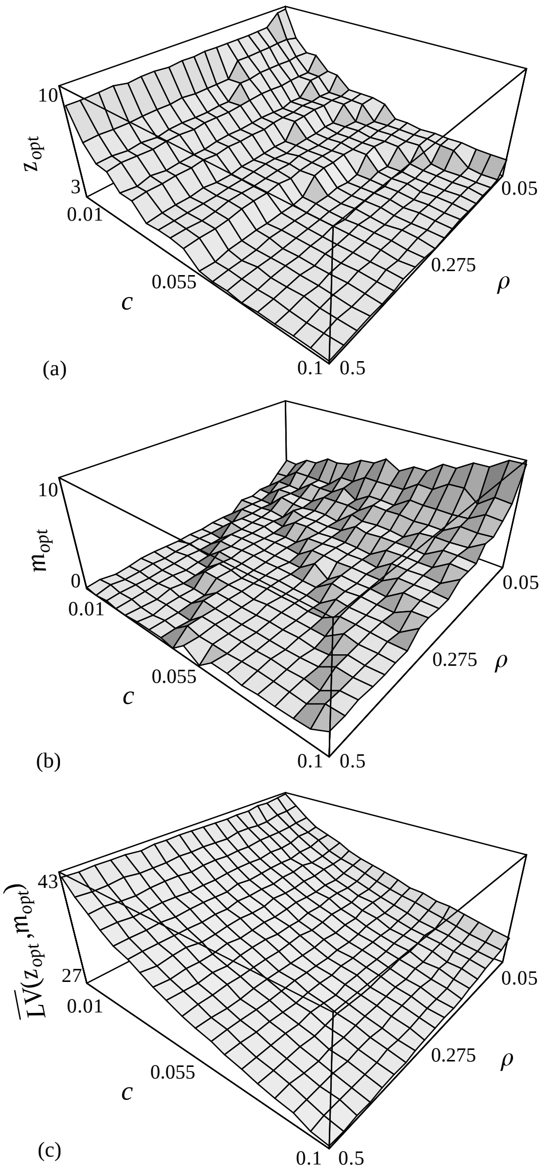


FIG. 5. (a) Optimal z, (b) optimal m, and (c) maximum expected lifetime payoff ($\overline{LV}(z_{opt}, m_{opt})$) are plotted as a function of ρ and c. See text for details.

likely to gain a benefit from memory. That is, the likelihood of achieving $\bar{V}_{\mathbf{C}}(z)$ over $\bar{V}_{\mathbf{I}}(z')$ increases as ρ decreases; thus, the smaller the value of ρ, the greater is the tendency for m_{opt} to increase as z increases (see Fig. 3).

2.4.4. *Optimal* (z, m) *Combinations*

Here we assume that sampling size and memory are free to vary independently of one another and seek the optimal (z, m) combination by searching numerically for the coordinates $(z, m) \in \mathbb{Z}^+ \times \mathbb{Z}^+$ that maximize the function in eqn (7). The results are shown in Fig. 5. In the figure, the three surfaces give the optimal z coordinate, the optimal m coordinate and the maximum expected lifetime payoff, respectively, for a given (ρ, c) combination.

As expected, if the cost of sampling (c) increases, the optimal sampling size decreases. As c increases, optimal memory also has a tendency to increase. In this case, remembering behaviors serves as a way to avoid the costly sampling process. When c goes up, $\bar{V}_{\mathbf{S}}(z)$ in Fig. 4 decreases for all z values greater than one while $\bar{V}_{\mathbf{C}}(z)$ and $\bar{V}_{\mathbf{I}}(z)$ remain unchanged. Thus, the potential benefit of using memory increases, while the potential cost of using memory decreases (see Fig. 4); consequently optimal memory size should increase with increasing c. As the stimuli become more reliable (as ρ decreases), the optimal memory size increases. Also, the optimal sampling size has a tendency to increase as ρ decreases. The idea here is that as stimuli become more reliable it pays to invest more into obtaining information (higher z) and then spreading the cost over a longer memory (higher m) since the memories have a higher chance of being consistent. The maximum expected lifetime payoff decreases as stimuli become less reliable or the cost of sampling increases, that is, as ρ or c increases.

3. Payoff Landscapes: Simulation Model and Results

3.1. SIMULATION DESCRIPTION

In order to check the analytical predictions and to explore scenarios not amenable to mathematical analysis, we conducted an agent-based simulation. In the simulation, the organism interacts with a randomly chosen stimulus each time step. Each stimulus is associated with a mapping between behaviors and payoffs. If the stimulus is novel or forgotten, the organism samples from its behavioral repertoire. Otherwise, the organism reuses a behavior from memory. Every time step there is a chance that any stimulus will change—such change is achieved by a random permutation of the behaviors on the payoffs. Table 1 gives the values or range of values for all parameters and variables used in the simulation study.

The simulation results can be compared directly with analytic results. In Fig. 6, the expected lifetime payoffs computed using eqns (7)-(14) are compared with averaged lifetime payoffs of 100 000 simulated individuals. In the case shown, and other cases considered but not shown, the fit between the mathematical and simulated data is quite good.

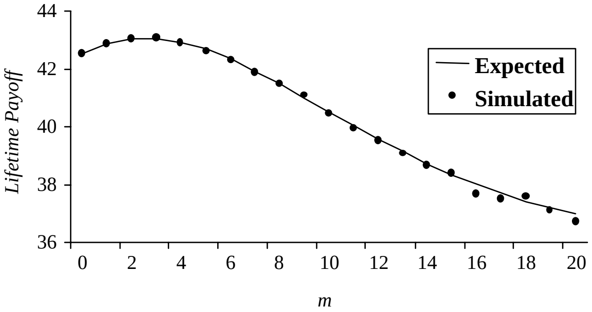


FIG. 6. Using eqns (7)–(14), we computed the expected lifetime payoff for the following parameter settings: $\rho = 0.05$, $c = 0.01$, $n = 10$, $N = 10$ and $T = 50$. The variable z was set to 6. Values for lifetime payoff were computed and simulated for memory values $0 \leqslant m \leqslant 20$. The simulated points were generated by averaging the lifetime payoffs of 100 000 individuals behaving in the environment described by the above parameter set. The fit between simulation and computation is quite good.

TABLE 1
Values and ranges of values of variables, parameters and sets used in the simulations

	Description	Values for simulation
Variable		
z	Number of behaviors sampled	1–15
m	Number of time steps a memory is retained	0–20
s	Number of stimuli observed per time step	1–15
Parameter		
n	Behavioral repertoire size.	10
N	Number of stimuli	10
T	Number of time steps per lifetime	50
c	Slope of the cost of sampling line	0.01–0.05
ρ	Probability of stimulus change per time step	0.0–1.0
Set		
B	The collection of n behaviors ($\beta_i \in$ B)	(arbitrary labels)
Π	The collection of n payoffs ($\pi_i \in \Pi$)	$\{0, 1/9, 2/9, 3/9, \ldots, 1\}$
$\sum$	The collection of N stimuli ($\sigma_i \in \sum$)	(arbitrary labels)

3.2. THE EFFECT OF SENSORY CAPABILITY

3.2.1. *Learning with Extended Sensory Capability*

As Dukas (1998a) discusses, an animal not only learns specific behaviors corresponding to specific stimuli, but may also be capable of selecting the stimulus to which it responds. Thus, an animal may sort available stimuli by past success and respond selectively to the "best" one. We incorporate this feature into our simulation by allowing the organism to respond to s (picked randomly with replacement) stimuli each time step ($s \geqslant 1$). The variable s serves as a proxy for sensory capability. We assume there is no explicit cost to sensory capability. If all observed stimuli are unfamiliar (i.e., novel or forgotten), the organism samples in response to one of the stimuli. As before, sampling entails some cost.

The organism reuses a behavior if at least one of the observed stimuli exists in the organism's long-term memory. Moreover, if the number of stimuli observed and remembered is greater than one, the organism picks the "best stimulus," that is, the stimulus whose corresponding remembered behavior yielded the highest payoff in the past. Consequently, the organism now needs to record three commodities into its long-term memory: the stimulus to which it responds, the behavior it employs and the payoff achieved for that behavior. If a behavior is reused from memory, the stimulus and behavior are refreshed in memory with the value representing the *current* payoff of the behavior. As before, when memory is used, the sampling cost is avoided.

With $s > 1$, simulations are used to generate average payoff surfaces to approximate $\overline{LV}(z, m, s)$. We focus on the effects of sampling cost (c) and stimulus reliability (ρ) on the optimal cognitive coordinates.

3.2.2. *Optimal Learning with Sensory Capability*

We consider the following ranges for the cognitive variables: $1 \leqslant z \leqslant 15$, $0 \leqslant m \leqslant 10$, and $1 \leqslant s \leqslant 15$. For each cognitive triplet (z, m, s), the lifetime payoff is obtained via simulation and averaged over 100 000 individuals. The optimal coordinates (z_{opt}, m_{opt}, s_{opt}) correspond to a maximum average lifetime fitness for the full landscape. In Fig. 7, we give the optimal

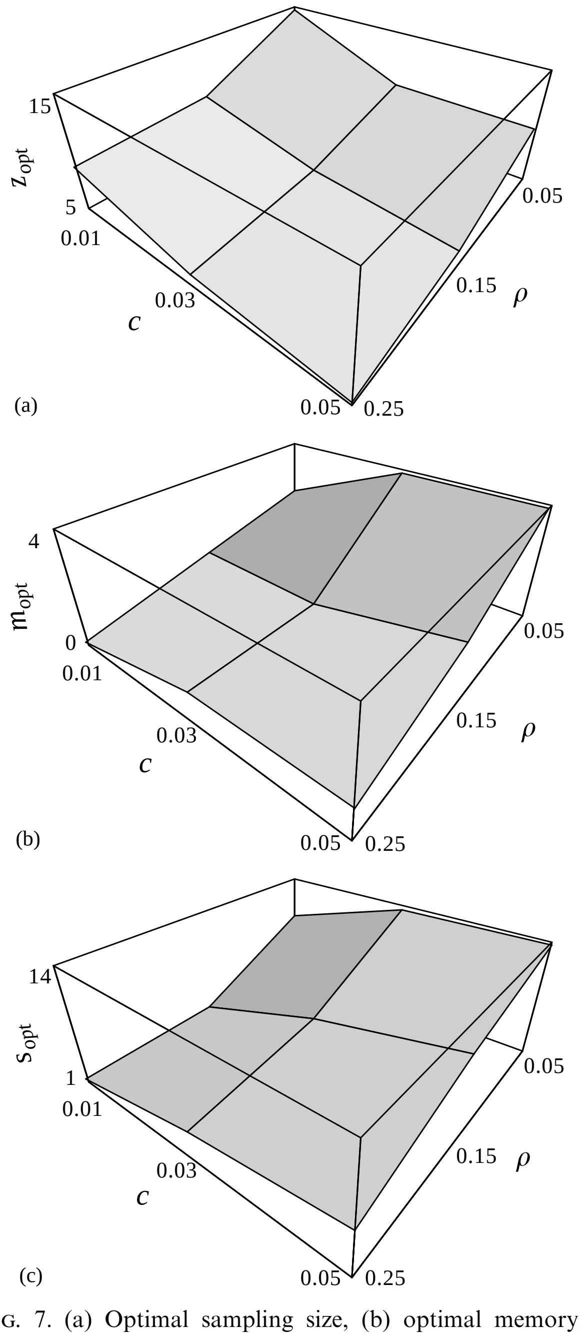


FIG. 7. (a) Optimal sampling size, (b) optimal memory, and (c) optimal sensory capability are shown as functions of ρ and c. See text for details. Note that when $m_{opt} = 0$, the value for s_{opt} could be any value in its range. This is because without memory, any value for sensory capability gives an identical strategy: the organism always samples from a randomly chosen stimulus from the set of observed stimuli (thus, the number of stimuli observed does not matter). So, in part (c), the value of s_{opt} when $\rho = 0.25$ and $c = 0.01$ could have been any value between 1 and 15.

coordinates. The general conclusions are that as c increases, z_{opt} tends to decrease while both m_{opt} and s_{opt} tend to increase. As ρ decreases z_{opt}, m_{opt} and s_{opt} all tend to increase. Thus, it appears that optimal sensory capability and optimal memory behave similarly in response to c and ρ.

To see the connection between memory and sensory capability, imagine that an organism samples behaviors in response to an unfamiliar stimulus. How many time steps are expected to pass before the organism revisits this specific stimulus? As s increases, the expected time of return to a stimulus decreases. If τ is the time of return to a specific stimulus, the expectation and variance of τ are as follows:

$$\bar{\tau} = \frac{1}{1 - ((N-1)/N)^s} \tag{16}$$

and

$$var(\tau) = \frac{((N-1)/N)^s}{(1 - (N - 1/N)^s)^2}. \tag{17}$$

From eqns (16) and (17) it is clear that $\bar{\tau}$ and $var(\tau)$ are decreasing functions of s. If an organism has memory length $m>0$ and if it records a stimulus–behavior–payoff triplet into its memory at time step t, it will hold on to that information for either m time steps or until it uses it to respond to the same stimulus at some time step before $t+m$. What eqns (16) and (17) suggest is that the probability that the organism will use its memory increases with sensory capability. Thus, as stimuli become more reliable (as ρ decreases in Fig. 7), the probability of memory use increases due to a longer memory (larger m) and lower return time to stimuli in memory (larger s).

In this section, we have only considered cases where the stimulus reliability is homogenous (i.e., $\rho_\sigma = \rho$ for all $\sigma \in \{\sigma_1, \sigma_2, \sigma_3, \ldots, \sigma_N\}$). Thus, all stimuli are equivalent and sensory capability simply enhances the probability of remembering the first several stimuli encountered. However, the model extension in Section 3.2.1 was framed to allow the organism to pick the *best stimulus*. When stimuli are homogeneous, there exists no best stimulus and the organism will choose a stimulus (from the s observed stimuli) based on chance (i.e., stochastic results of sampling). In order for sensory capability and memory to filter out the best stimulus, there must be something to filter—this leads to a consideration of heterogeneity in stimulus reliability.

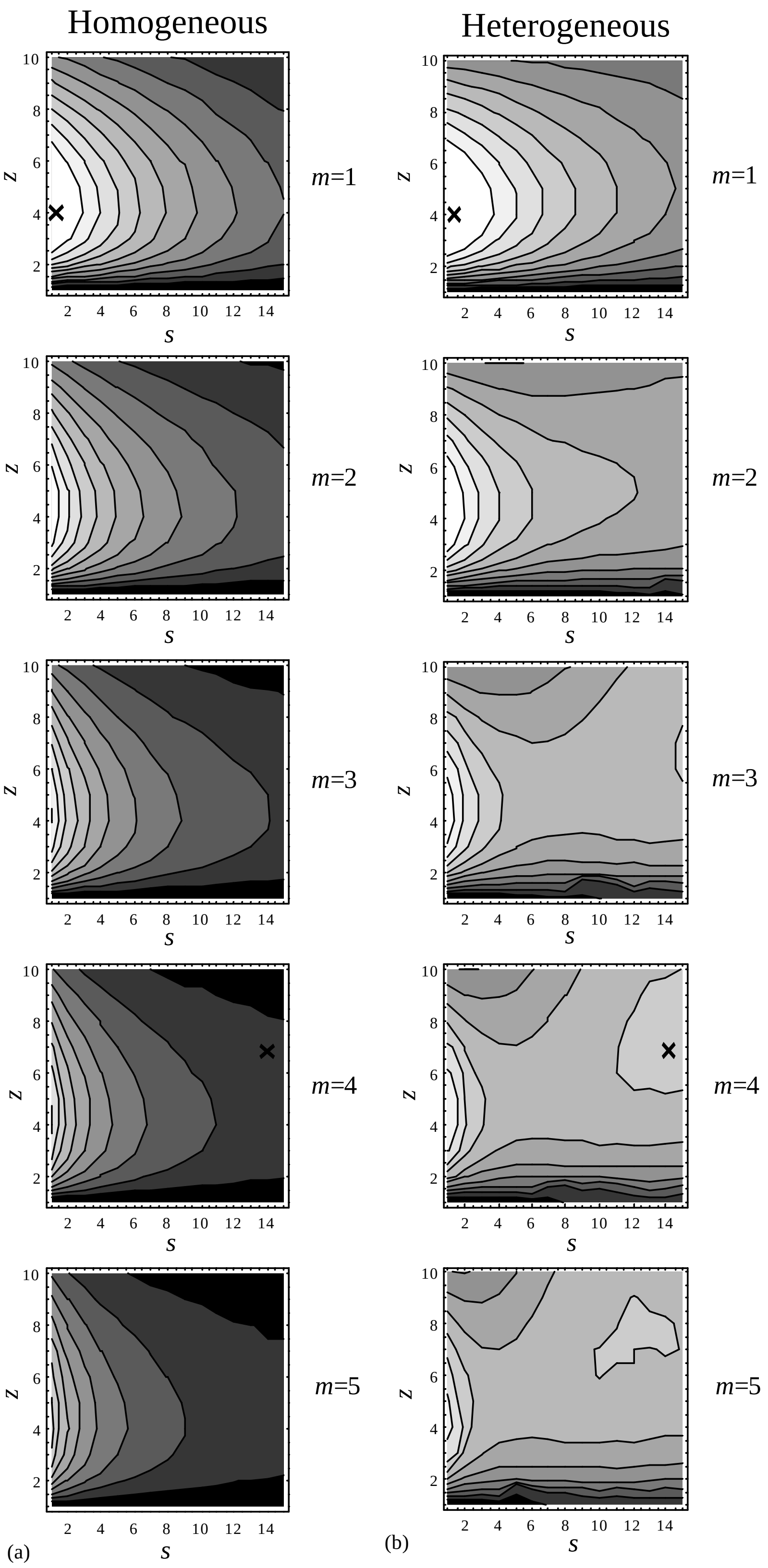
3.3. STIMULUS HETEROGENEITY

In what follows, we will restrict our attention to individuals with memory only; i.e., $m>0$. In some cases, $m=0$ will produce a higher payoff value and we will mention such cases as they arise.

Consider $N=10$ stimuli, all of which are completely unreliable ($\rho = 1$ for all stimuli). That is, the payoffs of the $n=10$ behaviors undergo a random permutation every time step. The organism lives for $T=50$ time steps and the cost of sampling is determined by $c=0.04$. We search for peaks on the average payoff landscape over all (z, m, s) combinations where $1 \leqslant z \leqslant 15$, $1 \leqslant m \leqslant 10$, and $1 \leqslant s \leqslant 15$ (see Appendix A for a description of the peak-isolating algorithm). We find a single local maximum at $(z_{opt}, m_{opt}, s_{opt}) = (4, 1, 1)$ [see Fig. 8(a)]. Since everything is unreliable, the organism does best by minimizing its probability of remembering anything at all ($m=1$ and $s=1$). In fact, if $m=0$, the average lifetime payoff is even higher and the organism benefits by losing memory altogether. An organism employing this peak strategy samples during most of its life (if $m=0$, it samples all of its life)

⟶

FIG. 8. (a) The column of contour graphs is part of the payoff landscape with all stimuli unreliable (i.e., homogeneous reliability). Each contour is a hyperplane with m held at some constant value (m = 1,2,3, 4, and 5 in the figure). In the contours, dark regions indicate low average lifetime payoff and light regions indicate high average lifetime payoff. In the homogeneous environment, there is a single peak in the payoff landscape at ($z=4$, $m=1$, $s=1$), which is marked with an X. (b) The column of contour graphs is part of the payoff landscape with one stimulus reliable and all the other stimuli unreliable (i.e., heterogeneous reliability). There are two peaks in the payoff landscape, one at ($z=4$, $m=1$, $s=1$), which we call the "roving peak," and the other at ($z=7$, $m=4$, $s=14$), which we call the "homing peak." Both peaks are marked with X's.


Homogeneous
Heterogeneous
m=1
m=2
m=3
m=4
m=5
z
s
2 4 6 8 10
2 4 6 8 10 12 14
(a)
(b)

and the sampling size $z = 4$ maximizes its expected return per sampling bout [solving eqn(15) for z with $n = 10$, $\rho = 1$, $m = 1$ and $c = 0.04$ gives $z_{opt} \approx 4.19$].

Now imagine an environment identical to that described above with the exception that one stimulus (say, stimulus 1) is completely reliable ($\rho_{\sigma_1} = 0$, while $\rho_\sigma = 1$ for $\sigma \in \{\sigma_2, \sigma_3, \sigma_4, \ldots, \sigma_{10}\}$). Thus, we have *heterogeneity* in stimulus reliability in this new environment. When we perform the peak-isolating algorithm (see Appendix A) on this slightly altered environment, we discover two local peaks! One of the peaks is the same as in the homogenous environment–namely, (4,1,1). However, we also find a peak in the neighborhood of (7,4,14). Note that these two peaks differ in all the cognitive coordinates. Figure 8(b) presents the landscape. We label the (4,1,1) peak the "roving strategy" and the (7,4,14) peak the "homing strategy" for reasons that will become clear below.

The roving strategy is essentially a sampling strategy. It is characterized by a low memory ($m = 0$ gives the highest average lifetime fitness), a low sensory capability ($s = 1$ if $m > 0$) and a relatively low sampling size ($z = 4$ in this case). An organism using such a strategy records little or no information to guide future behavior. Rather, it "takes each day as it comes," sampling in response to whichever stimulus turns up. Consequently, the fraction of visits that each stimulus receives (taken over 5 000 000 time steps) has essentially a uniform distribution [see Fig. 9(a)]. Stimulus 1, which is completely reliable, is visited no more often than the other unreliable stimuli. In this way, the organism simply wanders from stimulus to stimulus.

The reliable stimulus in this heterogeneous environment can be likened to "a needle in a haystack." An organism employing the roving strategy does not pursue the needle and does just fine via sampling. The homing strategy, on the other hand, involves filtering the needle from the hay. A homing individual possesses a higher memory ($m = 4$), a much higher sensory capability ($s = 14$) and a higher sampling size ($z = 7$). We suggest that this strategy involves picking out the reliable stimulus and then revisiting this stimulus as much as possible. We see that the majority of visits of the homing type occur at the reliable stimulus 1 [see Fig. 9(b)].

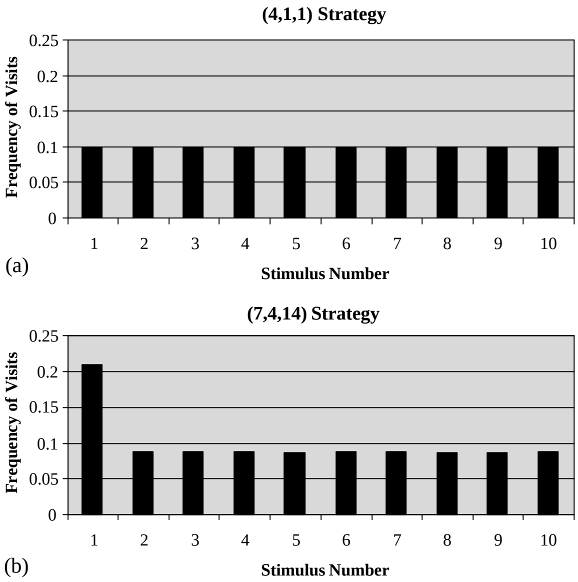


FIG. 9. The frequency of visits to each stimulus in the heterogeneous environment described in section 3.3 averaged over 5×10^6 time steps. (a) The roving strategy ($z = 4$, $m = 1$, $s = 1$) shows a uniform distribution for stimulus visits. (b) The homing strategy ($z = 7$, $m = 4$, $s = 14$) shows that visits to stimulus 1 (the reliable stimulus) are more than twice as likely as visits to any other stimulus.

We call this the homing strategy because the organism, which may potentially utilize any of the stimuli, specializes on a certain stimulus, visiting it at a higher frequency than all the others. This is not due to some pre-programmed preference that has the organism hunting or waiting for stimulus 1. Rather, given a very simple set of rules (sample when stimuli are unfamiliar, employ best remembered behavior when familiar stimuli are experienced) the specialization emerges under the cognitive combination $z = 7$, $m = 4$, $s = 14$ in the heterogeneous environment outlined above.

We turn now to why there is a valley between the two peaks. Simply, what is good for one strategy decreases the effectiveness of the other. Giving a roving type more memory (or sensory capability) is detrimental since the organism will remember behaviors for the unreliable stimuli most of the time. Thus, most of the time the roving type will suffer from its memory. Increase the memory and sensory capability enough and we cross the valley into the homing domain. Here decreases in memory and sensory capability will compromise the strategy. Decreasing memory means the homing type may not hold on to the reliable stimulus long enough to compare its consistent (and, on average, high) payoffs with the inconsistent payoffs of the unreliable stimuli, which will affect its ability to sort the reliable stimulus from the unreliable stimuli. Decreasing the sensory capability may mean that the homing individual does not experience the reliable stimulus often enough to guarantee its residence in memory [see eqns (16) and (17)].

This landscape shows that there are two locally optimal ways to process information and interact with stimuli in the same environment. The relative heights of the peaks will change with different parameter settings and different fractions of reliable stimuli, but the point is that using the same rules, organisms may proceed to interact with the same environment in very different ways with comparable success.

4. Discussion

4.1. ENVIRONMENT AND COGNITION

4.1.1. *Environmental Reliability*

Several theoretical investigations have shown that it may be best to forget (or heavily discount) past information when environmental parameters, which are estimated by the organism through experience, are uncertain and changing (McNamara & Houston, 1987; Mangel, 1990). In these cases, as well as in our model, no explicit cost of memory is assumed. Rather, in variable environments, there are *implicit* costs to memory —specifically, the cost of reusing information that is no longer appropriate. Memory is valuable only when it is reliable and such reliability will depend on rates of environmental change. Abel *et al.* (1998) describe genes whose products may function as regulators that prevent memory storage. In analogy with tumor suppressors, these "memory suppressor genes may decrease synaptic strength in much the same way that tumor suppressor genes stop or limit growth" (Abel *et al.*, 1998, p. 279). How these genes function within the network of other genes relevant to information storage should depend intimately on the value of memory *in the context of the organism's environment*.

In an environment where all the stimuli are reliable, memory is bound to be valuable (however, fitness may show diminishing returns with memory increase in a constant environment, e.g., Shafir & Roughgarden, 1996). If the stimuli become less reliable, a shorter memory length may be adaptive. Environmental reliability may play a key role in determining the value of memory in learning organisms.

Such reliability is the focus in Stephens' model for the evolution of learning (Stephens, 1991). Stephens shows that the predictability needed to favor learning must operate *within* the generations; however, the requisite variability needed to disfavor a fixed (i.e., non-learning) strategy can operate either within or between generations (see also Arnold, 1978; Stephens, 1987). Although we do not consider multiple generations, we see that "too much variability" is deleterious to learning as we define it. Stephens' "within generation predictability" is analogous to the reliability parameter in our model. In agreement with Stephens, we find that when all the stimuli are unreliable, recording behaviors into memory for future use is not optimal (i.e., when $\rho = 1$, $m = 0$ is the optimal memory size). That is, if stimuli are changing every time step within a generation, using memory to guide behavior loses all value.

Environmental reliability has effects on other cognitive variables as well. The results of Section 2.4.4 show that optimal sampling size increases with increasing stimulus reliability. This is a memory-mediated effect, since in more reliable environments, optimal memory length will be higher. If the stimuli are reliable, then investing more in remembering a *better* behavior (i.e., taking a larger sample) will be the optimal strategy (see Section 2.4.2). Although the cost of sampling increases with z, this cost is not incurred when the organism remembers. The organism "pays off" this extra cost by using *consistent* behaviors from memory a certain fraction of the time steps and avoiding any cost of sampling for these time steps. Sensory capability will also increase with increasing reliability (see Section 3.2). We conjecture that sensing more stimuli each time step increases the likelihood of using memory, which is favored in a reliable world. Thus, the value of the entire learning strategy is influenced by environmental reliability.

4.1.2. *Cognition in Heterogeneous Environments*

So far, we have discussed reliability as if it were a single entity. However, in an environment with many stimuli, differences in the reliability of stimuli are expected. In such a heterogeneous environment, the prediction of the Goldilocks principle that either learning *or* no learning occurs, is replaced by the prediction that there can be value in both learning *and* not learning. Specifically, there may be value in employing either a homing or a roving strategy. These strategies are interesting because they produce local optima in a payoff landscape. In homogeneous environments, there is always a single peak in the payoff landscape. Heterogeneity is necessary in order to "wrinkle the landscape." With stimuli differing in reliability, there exists a *pattern* in the environment. The organism can attend to the pattern (home) or ignore it (rove). But there is no "half-way" strategy in the heterogeneous environments we consider —peaks are separated by a distinct valley. These results suggest that heterogeneity may be important in the maintenance of different types of information processing strategies where similar resources (stimuli) are exploited.

4.1.3. *Rules of Thumb*

For some time there has been emphasis on the importance of "rules of thumb" in producing complex behavior (see Stephens & Krebs, 1986). The homing peak was so called because, on average, the organism using the associated strategy would specialize on reliable stimuli. Observing this phenomenon, one might conclude that the organism is directly computing a derivative (rates of change of different stimuli) and then visiting the more reliable stimuli. While we do not deny this possibility, in our model, organisms do not keep explicit information on the rates of change of the stimuli they experience. The "derivative is computed" as a *by-product* of the comparison between memories of rewards.

4.2. LIMITATIONS AND EXTENSIONS

Our model tackles learning in a very simple and abstract fashion. There are several pieces missing from the picture and several unrealistic assumptions. First, we have assumed an extreme version of Thorndike's Law of Effect (see Peterson, 1991). That is, before sampling in response to an unfamiliar stimulus, the organism has equal probability of using any of its n behaviors. Experience alters the probabilities of behaviors in the most extreme way, so that a specific behavior (the most valuable of those sampled) is certain after the sampling bout. There are other, less dramatic, ways to update these probabilities. One general technique used in early mathematical psychology involves linear operators (Bush & Mosteller, 1955). Another extreme assumption concerns the "sliding window" form of memory. There are other, more realistic, ways to model memory. For instance, we might allow information in memory to be accessed with some probability, which is sensitive to the time of the last use of that information as well as the general intensity of its use.

Second, we have assumed that behaviors are uniformly distributed with respect to payoff and that all stimuli are identical in their payoff distributions. However, it is certainly reasonable to consider different payoff distributions, such as many low payoff behaviors with a few high payoff behaviors (a "difficult" stimulus) or many high payoff behaviors with a few low payoff behaviors (an "easy" stimulus). Besides differing in payoff distributions, stimuli might also differ in the *range* of potential payoffs, such that some stimuli might have a high upper bound in payoff ("high quality") while other stimuli have a low upper bound ("low quality"). It would be interesting to repeat the above analysis to investigate whether heterogeneity in payoff distribution ("difficulty") or payoff range ("quality") has effects comparable to heterogeneity in stimulus reliability.

Third, there may be explicit costs and/or constraints to the cognitive variables in the model. For instance, there may be limitations to how well an organism can assess the payoffs to certain behaviors during a sampling bout (in fact, the accuracy of payoff assessment may itself depend on time and effort spent with the relevant stimuli). Or consider sampling size. We have assumed that short-term memory is large enough to hold all the payoffs of the behaviors in the sampling bouts. However, in studies on humans, monkeys and pigeons, short-term memory appears to be limited (Dukas, 1998b), which, in our model, may constrain the sample size. Also, Dukas (1998b) discusses possible costs to long-term memory, such as extensive resource expenditure. Such a cost could be integrated into our model. Also there has been laboratory and field work done on memory interference (Lewis, 1986; Dukas, 1998b)—where the use of one piece of information interferes with the use of a second piece of information. In our model, interference would manifest itself as a complication in recalling a behavioral response to one stimulus when other stimulus–behavior pairs have been recorded.

Fourth, the variables investigated may influence the values of other model parameters or the structure of the model. For instance, sensory capability is measured as the number of stimuli the organism can process in a time step. However, as the *true* sensory capability of an organism is extended, the organism may increase the number of potential stimuli to which it reacts; that is, N might increase. Furthermore, with an extended sensory capability, the organism might be able to distinguish between two objects that were previously perceived as the *same* stimulus. If these two objects gave different payoffs for the same behaviors, the organism, by mistaking them for the same stimulus, would see this "one" stimulus as unreliable. However, if it could distinguish between the two objects, then the organism might improve the reliability of the stimuli in its environment; that is, ρ might decrease. As another example, consider spatial memory. If the organism can choose places to forage, nest, mate, etc. based on past experience, then the assumption that stimuli are chosen randomly with replacement each time step will be misplaced.

Fifth, we have omitted organismal variables from the model that affect cognition. Mangel (1993) discusses how motivation (some measure of the physiological state of an animal) may influence an animal's decisions. For instance, how many eggs a female insect carries may influence her choice to accept a host plant species for oviposition (Mangel, 1993). Here, we do not consider the physiological state of our model organism. Mangel discusses ways that learning and motivation may be integrated to give more comprehensive models of animal behavior and decision-making.

Sixth, we have omitted variables from the environment that will inevitably influence the optimal learning strategy. For instance, in many dynamic programming models, risk of predation is explicitly incorporated into the model (see Mangel & Clark, 1988; Houston & McNamara, 1999) and will influence the optimal sequence of behaviors. In the context of our model, if sampling is costly not only in terms of time, but also in terms of exposure to predators (whereas employment of a learned behavior is "safer") then the optimal learning strategy would certainly be affected. The incorporation of mortality is bound to affect the value of other cognitive parameters as well, such as memory and sensory capability. Because a learned behavior can never be valuable after the organism dies, mechanisms by which the return time to stimuli is reduced (e.g., higher sensory capability) may be favored in an environment characterized by higher mortality. To investigate such possibilities, we would have to include the chance of mortality for our learner at every time step.

Seventh, the "rules of learning" used in this model may not be appropriate for many organisms. In Appendix B we discuss ways to improve the rules. Ultimately, however, modeling the learning process should be informed by the natural systems themselves. Given our simple rules, it is intriguing to see distinct strategies emerge corresponding to pure sampling and standard specialization under heterogeneous environments. In nearly all natural systems, different animals possess different ways to process information in their environment—the results from our model suggest that environmental heterogeneity may be important in the evolution or maintenance of this cognitive diversity.

Lastly, we have only investigated the manner in which the value of the organism's learning strategy is affected by the environment. We have neglected the effect of learning process on the state of the environment. As insects learn to handle certain flowers, they drain nectar reserves. As predators become adept at catching a particular prey item, the number of that item declines. Organisms not only respond to a stimulus, but also may change the stimulus in the process. The idea that organisms modify their environment and consequently their own selective pressures has been termed "niche construction" (see Lewontin, 1978,1982,1983; Odling-smee *et al.*, 1996; Laland *et al.*, 1996,1999). In Cohen's (1991) model, there was explicit consideration of the effect of the organism on the state of the environment (variance in patch quality).

Within the context of our model, the organism, through its actions, might influence the future payoff structure of the stimulus to which it reacts. The actions of organisms might either amplify or diminish heterogeneity in various properties of their stimuli. In order to explore these issues, a dynamic model would be ideal, where both the state of the learning population (for instance, the frequencies of different learning strategies) as well as the state of the stimulus population (for instance, the frequencies of given payoff structures and reliabilities) form the variables of the system. With such model extensions, cognition is no longer viewed as simply a response to some exogenous environment, but as the result of an evolutionary dialogue between the learning organism and the affected environment.

We thank Gabriel Kerr, Susan Ptak, Noah Rosenberg, Mark Tanaka, and Jeanne Wald for helpful discussions. Lauren Ancel, Ammon Corl, Aaron Hirsh, Jeremy Kendal, Henrique Pereira, and Karen Walag read previous versions of the manuscript and offered many insightful and thought-provoking points. This work benefited tremendously from the comments of Rustom Antia, Aviv Bergman, Carl Bergstrom, Peter Godfrey-Smith, Deborah Gordon, Kevin Laland, John Odling-Smee, Guy Sella, and five anonymous reviewers. We are indebted to Ammon Corl for his active participation in the initial formulation of the simulation. This research was supported in part by a NSF graduate fellowship to BK and NIH grant GM28016 to MWF.

REFERENCES

Abel, T., Martin, K. C., Bartsch, D. & Kandel, E. R. (1998). Memory suppressor genes: Inhibitory constraints on the storage of long-term memory. *Science* **279,** 338–341.

Ancel, L. W. (1999). A quantitative model of the Simpson-Baldwin effect. *J. theor. Biol.* **196,** 197-209. doi: 10. 1006/jtbi. 1998. 0833.

Arnold, S. J. (1978). The evolution of a special class of modifiable behaviors in relation to environmental pattern. *Am. Nat.* **112,** 415–427.

Beardon, A. F. (1996). Sums of powers of integers. *Am. Math. Monthly* **103,** 201–213.

Bergman, A. & Feldman, M. W. (1995). On the evolution of learning: representation of a stochastic environment. *Theor. Popul. Biol.* **48,** 251–276.

Bush, R. R. & Mosteller, F. (1955). *Stochastic Models for Learning*. New York: Wiley.

Cohen, D. (1991). The equilibrium distribution of optimal search and sampling effort of foraging animals in patchy environments. In: *Adaptation in Stochastic Environments* (Yoshimura, J. & Clark, C. W., eds), pp. 173–191. Berlin: Springer-Verlag.

Dukas, R. (1998a). Evolutionary ecology of learning. In: *Cognitive Ecology: The Evolutionary Ecology of Information Processing and Decision Making* (Dukas, R., ed.), pp. 129–174. Chicago: University of Chicago Press.

Dukas, R. (1998b). Constraints on information processing and their effects on behavior. In: *Cognitive Ecology: The Evolutionary Ecology of Information Processing and Decision Making* (Dukas, R., ed.), pp. 89–127. Chicago: University of Chicago Press.

Feldman, M. W., Aoki, K. & Kumm, J. (1996). Individual versus social learning: Evolutionary analysis in a fluctuating environment. *Anthropol. Sci.* **104,** 209–232.

Godfrey-Smith, P. (1998). *Complexity and the Function of Mind in Nature*. Cambridge: Cambridge University Press.

Harley, C. B. & Maynard Smith, J. (1983). Learning: An evolutionary approach. *Trends Neuro Sci.* **6,** 204–204.

Houston, A. I. & McNamara, J. M. (1999). *Models of Adaptive Behaviour: an Approach Based on State*. Cambridge: Cambridge University Press.

Johnston, T. D. (1982). Selective costs and benefits in the evolution of learning. *Adv. Study Behav.* **12,** 65–106.

Krebs, J. R., Kacelnick, A. & Taylor, P. (1978). Test of optimal sampling by foraging great tits. *Nature* **275,** 27–31.

Laland, K. N., Odling-Smee, F. J. & Feldman, M. W. (1996). The evolutionary consequences of niche construction: a theoretical investigation using 2-locus theory. *J. Evol. Biol.* **9,** 293–316.

Laland, K. N., Odling-Smee, F. J. & Feldman, M. W. (1999). Evolutionary consequences of niche construction and their implications for ecology. *Proc. Natl. Acad. Sci. U.S.A.* **96,** 10 242–10 247.

Lewis, A. C. (1986). Memory constraints and flower choice in Pieris-Rapae. *Science* **232,** 863–865.

Lewis, A. C. (1989). Flower visit consistency in Pieris-Rapae, the cabbage butterfly. *J. Anim. Ecol.* **58,** 1–13.

Lewontin, R. C. (1978). Adaptation. *Sci. Am.* **239,** 212.

Lewontin, R. C. (1982). Organism and environment. In: *Learning, Development and Culture* (Plotkin, H. C., ed.), pp. 151–170. New York: Wiley.

Lewontin, R. C. (1983). Gene, organism and environment. In: *Evolution from Molecules to Men* (Bendall, D. S., ed.), pp. 273–285. Cambridge: Cambridge University Press.

MANGEL, M. (1990). Dynamic information in uncertain and changing worlds. *J. theor. Biol.* **146,** 317–332.

MANGEL, M. (1993). Motivation, learning, and motivated learning. In: *Insect Learning: Ecology and Evolutionary Perspectives* (Papaj, D. R. & Lewis, A. C., eds), pp. 158–173. New York: Chapman and Hall.

MANGEL, M. & CLARK, C. W. (1988). *Dynamic Modeling in Behavioral Ecology*. Princeton: Princeton University Press.

MCNAMARA, J. M. & HOUSTON, A. I. (1987). Memory and the efficient use of information. *J. theor. Biol.* **125,** 385–395.

ODLING-SMEE, F. J., LALAND, K. N. & FELDMAN, M. W. (1996). Niche construction. *Am. Nat.* **147,** 641–648.

PETERSON, C. (1991). *Introduction to Psychology*. New York: Harper Collins.

PLOTKIN, H. C. & ODLING-SMEE, F. J. (1979). Learning, change and evolution: Inquiry into the telconomy of learning. *Adv. Study Behav.* **10,** 1–41.

SHAFIR, S. & ROUGHGARDEN, J. (1996). The effect of memory length on individual fitness in a lizard. In: *Adaptive Individuals in Evolving Populations: Models and Algorithms* (Belew, R. K. & Mitchell, M., eds), pp. 173–181. Reading, MA: Addison-Wesley.

STEPHENS, D. W. (1987). On economically tracking a variable environment. *Theor. Popul. Biol.* **32,** 15–25.

STEPHENS, D. W. (1991). Change, regularity, and value in the evolution of animal learning. *Behav. Ecol.* **2,** 77–89.

STEPHENS, D. W. & KREBS, J. R. (1986). *Foraging Theory*. Princeton: Princeton University Press.

Appendix A

Technique for Isolating Peaks

In this appendix, we describe the technique used to find local peaks in a payoff landscape. Due to the stochastic nature of the simulation, the payoff surfaces are somewhat bumpy. We have attempted to "smooth out" the landscape by computing the average lifetime payoff of a large number of individuals (100 000). However, simple hill climbing is bound to give some spurious peaks in the landscape due to stochastic anomalies. To minimize this effect, we employ the following procedure:

(I) For a given parameter set, simulate the average payoff landscape (as a function of z, m, and s) three times.

(II) For each simulated landscape, pick 100 random starting points (here z, m and s are picked uniformly from [1,15], [1,10], and [1,15], respectively).

(III) For each random starting point perform a simple hill climbing algorithm, where movement is always in the direction of the greatest lift in average lifetime payoff. This is done until a local maximum is attained.

(IV) Each of the 100 optimal coordinate triplets (producing local peaks) for a given landscape is compared to all 200 optimal triplets of the other two landscapes. If no optimal triplet in the other two landscapes is within a given neighborhood of the optimal triplet, it is discarded. Here the "neighborhood criterion" is defined as follows: the optimal triplet of a given landscape cannot be more than a Euclidean distance of 1 from the optimal triplet of another landscape. For example, if (2,5,12) corresponds to a local peak, then (3,5,12), (2,4,12), and (2,5,11) are all within its neighborhood, while (3,6,12), (2,5,10), and (7,1,9) are not.

(V) Those peaks that were reached five times or less (out of the 100) are ignored. The remaining peaks for a given landscape are inspected visually to confirm whether they appear as true peaks. See Fig. 8 for an example.

Appendix B

Expected Payoff Incorporating a Cutoff

Given that our model organism can only either sample or remember at each time step, is the set of rules we have prescribed optimal? Consider an organism that observes a single stimulus per time step ($s = 1$). Suppose this organism samples behaviors in response to an unreliable stimulus and from the behaviors sampled, selects one with the highest payoff. At some later time step, the organism reuses the behavior from memory, but now receives a low payoff (as the payoffs have permuted). If the organism keeps the new payoff value for the behavior in memory and revisits the same stimulus at some still later time step (before the memory has expired), we see a potential problem. Given the above rule set, the organism is forced to use a behavior that it remembers as having a low payoff value. In essence, the low payoff value itself gives the organism information (i.e., that the stimulus may be unreliable) and the above set of rules ignores such information. This shows that in certain cases, the above rule set is suboptimal.

While we will not pursue the optimal rule set in this paper, we can make some suggestions for improvement. To start, one might allow the organism to resample from a stimulus when it possesses information about the stimulus in memory. For instance, it might resample behaviors in response to a stimulus if the remembered behavior has a payoff less than or equal to some cutoff value w, which might be viewed as another cognitive variable, the payoff threshold for memory to be used. Note that the rule set we use above assumes $w<0$ (i.e., no resampling). However, for any (z, m, s) combination in a given environment, there will be an optimal w value that may be positive. When $s=1$, the procedure outlined in Section 2.4.1 can be used to deduce the optimal z, m, and w values in a given environment, as we now outline.

Let **S** be the event that an organism samples in response to an unfamiliar stimulus, **R** be the event that the organism resamples in response to a *familiar* stimulus (this occurs when the payoff of the *remembered* behavior is less than or equal to w), $\mathbf{I_A}$ be the event that an organism reuses an inconsistent behavior from memory with a *current* payoff above w, $\mathbf{I_B}$ be the event that an organism reuses an inconsistent behavior from memory with a current payoff below w, and **C** be the event that an organism reuses a consistent behavior from memory. The payoff of the remembered behavior here must be greater than w; otherwise, the organism resamples.

Here, we assume $w=(k^*-1)/(n-1)$, with $k^*\in\mathbb{Z}^+$ and $1<k^*<n$. The expected lifetime payoff is

$$\begin{aligned}\overline{LV}(z,m,w) = \sum_{t=1}^{T}\{&\bar{V}_\mathbf{S}(z, w)P_\mathbf{S}(t, m, w)\\&+\bar{V}_\mathbf{R}(z, w)P_\mathbf{R}(t, m, w)\\&+\bar{V}_\mathbf{C}(z, w)P_\mathbf{C}(t, m, w)\\&+\bar{V}_{\mathbf{I_B}}(z, w)P_{\mathbf{I_B}}(t, m, w)\\&+\bar{V}_{\mathbf{I_A}}(z, w)P_{\mathbf{I_A}}(t, m, w)\}\end{aligned}$$

with

$$\bar{V}_\mathbf{C}(z, w) = \frac{n^z(n-1)-(k^*)^z(k^*-1)-\sum_{k=k^*}^{n-1}k^z}{[n^z-(k^*)^z](n-1)},$$

$$\bar{V}_{\mathbf{I_B}}(z, w) = \frac{k^*-1}{2(n-1)},$$

$$\bar{V}_{\mathbf{I_A}}(z, w) = \frac{n(n-1)-k^*(k^*-1)}{2(n-k^*)(n-1)},$$

where again $k^*=w(n-1)+1$. $P_\mathbf{S}(t, m, w)$ is given by eqn (11) and $\bar{V}_\mathbf{S}(z, w)$ and $\bar{V}_\mathbf{R}(z, w)$ are given by eqn (4). For the probabilities of the remaining events, the following equation holds:

$$P_\mathbf{G}(t, m, w) = P_{\mathbf{G}|\mathbf{M}}(t, m, w)P_\mathbf{M}(t, m),$$

where $\mathbf{G}\in\{\mathbf{R}, \mathbf{I_A}, \mathbf{I_B}, \mathbf{C}\}$ and $P_\mathbf{M}(t, m)$ is given by eqn (10). The conditional probabilities are as follows:

$$\begin{aligned}P_{\mathbf{R}|\mathbf{M}}(t, m, w) = \sum_{i=1}^{\min(m,t-1)} &\Lambda(i, t, m)[(P_\mathbf{S}(t-i, m)\\&+P_\mathbf{R}(t-i, m, w))\left(\frac{k^*}{n}\right)^z\\&+P_{\mathbf{I_B}}(t-i, m, w)]\end{aligned}$$

$$\begin{aligned}P_{\mathbf{C}|\mathbf{M}}(t, m, w) = \sum_{i=1}^{\min(m,t-1)} &\Lambda(i, t, m)[((P_\mathbf{S}(t-i, m)\\&+P_\mathbf{R}(t-i, m, w))\left(1-\left(\frac{k^*}{n}\right)^z\right)\\&+P_\mathbf{C}(t-i, m, w)](1-\rho)^i\end{aligned}$$

$$\begin{aligned}P_{\mathbf{I_B}|\mathbf{M}}(t, m, w) = \sum_{i=1}^{\min(m,t-1)} &\{\Lambda(i, t, m)[((P_\mathbf{S}(t-i, m)\\&+P_\mathbf{R}(t-i, m, w))(1-(k^*/n)^z)\\&+(P_{\mathbf{I_A}}(t-i, m, w)\\&+P_\mathbf{C}(t-i, m, w))]\\&\times(k^*/n)(1-(1-\rho)^i)\},\end{aligned}$$

$$\begin{aligned}P_{\mathbf{I_A}|\mathbf{M}}(t, m, w) = \sum_{i=1}^{\min(m,t-1)} &\{\Lambda(i, t, m)[((P_\mathbf{S}(t-i, m)\\&+P_\mathbf{R}(t-i, m, w))(1-(k^*/n)^z)\\&+(P_{\mathbf{I_A}}(t-i, m, w)\\&+P_\mathbf{C}(t-i, m, w))]\\&\times(1-(k^*/n))(1-(1-\rho)^i)\\&+P_{\mathbf{I_A}}(t-i, m, w)(1-\rho)^i\}.\end{aligned}$$