Why some memories do not last a lifetime: dynamic long-term retrieval in changing environments

Aimee S. Dunlap,[a] Colleen M. McLinn,[b] Holly A. MacCormick,[c] Matthew E. Scott,[d] and Benjamin Kerr[e]
[a]Department of Ecology and Evolutionary Biology, University of Arizona, Box 210088, Tucson, AZ 85721, USA, [b]Cornell Laboratory of Ornithology, Cornell University, 159 Sapsucker Woods Road, Ithaca, NY 14850, USA, [c]Department of Ecology and Evolutionary Biology, University of California, Santa Cruz, Santa Cruz, CA 95064, USA, [d]St Croix Vineyards and Center for Complex Biological Systems, Stillwater, MN 55082, USA, and [e]Department of Biology, University of Washington, Box 351800, Seattle, WA 98195, USA

Memory is a fundamental component of learning, a process by which individuals alter their behavior through experience. Although memory most likely has explicit costs such as synaptic maintenance and metabolic demands, there are also implicit costs to memory, in particular, the use of information that is no longer appropriate or is incorrect. Specifically, the period of retrievability for memories, or "memory window," should be sensitive to the rate of environmental change of information stored in memory. Much empirical data suggest that memory length—this period of retrievability—changes with both the age and state of the individual. Here, we use a dynamic programming approach to examine how optimal memory retrieval might change within the lifetime of the individual learner. We find that optimal memory length varies with both age and state (e.g., energy reserves) of the organism and that features of the environment determine how this change in memory occurs. In our model, retrieval decreases as the environment becomes unreliable but roughly increases with the cost of living. Cost of living interacts with the state of the organism: with high cost of living, an organism in a very poor state should have a long memory length, but an organism in a very good state with low costs of living should have a short memory length. Finally, we find there are circumstances where it is optimal for memory retrieval to decline toward the end of the lifetime. Because this framework does not incorporate inevitable degradation of neural mechanisms, this result implies that memory loss with age might actually be adaptive. *Key words:* dynamic programming, environmental variability, learning, memory, optimality, stimulus reliability. *[Behav Ecol 20:1096–1105 (2009)]*

Learning is responsible for many of the flexible behaviors animals use when adjusting to changing environments. Ultimately, learning depends on an animal's ability to store and retrieve information about its world, that is, learning is tightly entwined with the memory of that information (Bouton 1994; Shettleworth 1998). Consequently, the value of memory often relates to the value of learning. On a simple level, learning is valuable when information stored in memory (e.g., a good response to a stimulus) remains useful over time (e.g., the same response to the same stimulus continues to be good). However, learning and memory lose value when the environment changes in ways that make learned responses unreliable. Indeed, possessing memory might be costly if it repeatedly leads an organism to employ inappropriate behaviors (e.g., if the environment changes extremely rapidly making remembered responses perpetually inappropriate). This is similar to the effects of ecological or evolutionary traps, where previously reliable information leads to a maladaptive outcome because the situation has changed (Schlaepfer et al. 2002). If memory is evolutionarily adaptive, one would expect that various properties of memory (encoding, consolidation, retrieval, etc.) would be tuned to the historical rate of environmental change. Several mathematical models have suggested

that the amount of environmental variability influences the evolutionary advantages of learning and memory (Arnold 1978; McNamara and Houston 1987; Stephens 1987; Mangel 1990; Cohen 1991; Stephens 1991; Bergman and Feldman 1995; Feldman et al. 1996; Kerr and Feldman 2003).

Given that memory is employed throughout the lifetime of an individual, should memory length, that time for which memories are retrievable, change within that lifetime in response to particular patterns of environmental variability? For simplicity and tractability, most learning models have incorporated memory as a fixed quantity (McNamara and Houston 1987; Mangel 1990; Shafir and Roughgarden 1996; Kerr and Feldman 2003), often as a "sliding window" of constant length (but see Smulders and Dhondt 1997; Hirvonen et al. 1999).

However, overwhelming evidence suggests that memory changes over the lifetimes of individuals. For instance, aspects of memory have been shown to decline with age in a large number of species, from humans to crustaceans (Tomsic et al. 1998; Davis et al. 2003; Punzo and Chavez 2003; Shukitt-Hale et al. 2004). In humans, this trend is present in healthy older adults—it is not attributable solely to the effects of Alzheimer's and/or dementia (e.g., Craik and Jennings 1992; Zelinski and Burnight 1997; Davis et al. 2001). One explanation for memory loss suggests that aged individuals acquire such a large number of memories that retrieval errors (Spear 1973; Squire 1989; Cohen 1996), perhaps caused by mechanisms such as memory interference, become more common (Bouton and Moody 2004; Mery and Kawecki 2004). Another explanation attributes memory loss to the inevitable decay of synaptic connections with age (Salthouse 1996). Poor memory

---

or learning abilities are also found at very young ages (Guo et al. 1996; Spreng et al. 2002). Known as infantile amnesia, this is often explained as the result of the developing brain (Neissen 2004). In addition to age, the physical state of the individual (such as its energy reserves or stress level) may influence learning and memory (McEwen and Sapolsky 1995; Guo et al. 1996; Xia et al. 1997; Yanai et al. 2004). For instance, whereas acute stress might actually improve memory in some cases, such as with posttraumatic stress disorder (McGaugh and Roozendaal 2002), chronic stress and increased cortisol levels can result in memory deficits (McEwen and Sapolsky 1995; Seeman et al. 1997; Lupien et al. 1998).

In the aforementioned explanations, memory loss results from the inevitable degradation or inherent constraints in retrieval caused by advanced age or prolonged stress. Another compelling possibility is that memory change might serve some evolutionarily adaptive role. Kraemer and Golding (1997) suggest that the primary mechanism for adaptive forgetting may be in the retrievability of information. Importantly, new evidence suggests that forgetting need not be a passive process (Anderson et al. 2004; Wylie et al. 2008).

To explore the possibility of evolutionarily adaptive forgetting, we use a dynamic programming approach—an established method for determining optimal strategies across lifetimes (Mangel and Clark 1988; Clark 1993; Houston and McNamara 1999). Within this modeling framework, we focus on the long-term retrieval of information, assuming that all experiences of the organism have been successfully encoded and consolidated. For effective memory, an organism needs both successful acquisition and retrieval. A failure at any point along the way results in forgetting. We do not consider mechanism, though in our framework forgetting may be due to an active process like suppression (Anderson et al. 2004), the decreasing availability of stored information over time, which Schacter calls transience (Schacter 1999), or due to a simple failure to retrieve the information (Loftus and Loftus 1980). We explore the effects of age, physical state, and environmental conditions on optimal retrieval. Our model suggests that even if the environment changes at a constant rate, optimal memory length may actually decline with age.

THE MODEL

Here we consider a learning organism that lives for a maximum of T discrete time steps, with its age at any point in time given by the variable t (the number of time steps lived so far). At each time step, there is a fixed positive probability of death, δ. This organism inhabits an environment with N stimuli to which it may respond. At every time step, the organism is exposed to one of the N stimuli (chosen at random with equal likelihood and independently of past exposures). If the stimulus is unfamiliar, the organism will need to spend some time learning how to respond appropriately to the stimulus. This learning process within a time step can be thought of as being on the timescale of short-term memory. For simplicity, we assume that the organism that learns about an unfamiliar stimulus always arrives at a behavior that yields a constant payoff, which we label π_u.

If the stimulus is familiar (i.e., remembered), then the organism must have interacted with it in a previous time step and remembers it from that time step. Here, we use a "window" model for memory retrieval—the organism can remember and retrieve information from the previous m time steps. If the organism remembers the behavior it employed previously for a stimulus and employs this behavior again when revisiting the familiar stimulus, there are 2 possible outcomes: 1) the appropriate response to the stimulus is the same as it was before and the organism receives a payoff π_c (a payoff for

a correct response) or 2) the appropriate response to the stimulus has changed and the organism receives a payoff of π_i (a payoff for an incorrect response). Again, for simplicity, we assume that the payoffs for appropriate or inappropriate remembered responses have constant values.

Thus, the expected payoff of a familiar stimulus depends on the rate of change of stimuli. Here, we assume that every stimulus changes independently with probability ρ every time step (as ρ increases, the stimuli become less reliable). We also assume $\pi_i < \pi_u < \pi_c$, as a correctly remembered behavior saves the sampling time spent on learning to respond appropriately to an unfamiliar stimulus (Lewis 1986). However, an incorrectly remembered behavior wastes more time than learning about a stimulus from scratch (e.g., time spent employing the now inappropriate behavior in addition to time spent learning the stimulus again). We assume that the payoffs are in terms of energy reserves, such that time spent learning about stimuli or time wasted employing incorrect behaviors results in a lower energy payoff for the time step. At any time, we let the energy reserves of our organism be x. We restrict x to a range of values, $0 \leq x \leq X$. We also assume that there is a cost of living, κ, imposed each time step (i.e., energy reserves removed from the organism due to the metabolic demands of living). The parameter κ is measured in the same units as x (e.g., calories), and we focus on cases in which the cost of living makes memory necessary for x to increase over a time step, that is, $\pi_i < \pi_u < \kappa < \pi_c$.

In the manuscript, we explore the idea that it may be advantageous for the length of the memory window, m, to vary as a function of the energy reserves of the organism, x, as well as its age, t (i.e., $m = m(x, t)$). Here, we use a dynamic programming approach to find the function that maximizes reproductive success of our model learner, $m^*(x, t)$, where the asterisk denotes the optimal memory. That is, we ask how the memory window might optimally change with state and age of the organism.

To this end, it helps to produce another function that measures maximal future expected reproduction. We call this function $F(x, t, T)$, the maximal future expected reproductive success of an organism at age t in state x (recall that T defines the maximum lifetime of an organism). Using F as a "common currency," the trick to dynamic programming involves writing a backward recursion in time. In Appendix A, we derive the following recursion,

$$F(x,t,T)$$

$$= (1-\delta) \max_{0 \leq m < t} \left\{ \begin{array}{l} \alpha^m F(u(x), t+1, T) \\ + \omega[1-\beta^m] F(c(x), t+1, T) \\ + (1-\alpha^m - \omega[1-\beta^m]) F(i(x), t+1, T) \end{array} \right\},$$

$$(1)$$

with

$$\alpha = \frac{N-1}{N}, \quad \beta = \frac{(N-1)(1-\rho)}{N}, \quad \omega = \frac{1-\rho}{1+(N-1)\rho}, \quad (2)$$

$$u(x) = \text{chop}(x + \pi_u - \kappa; 0, X), \quad (3)$$

$$c(x) = \text{chop}(x + \pi_c - \kappa; 0, X), \quad (4)$$

$$i(x) = \text{chop}(x + \pi_i - \kappa; 0, X). \quad (5)$$

The function chop simply keeps the state of the organism in the allowed range between 0 and X. Generally, chop$(x; a, b)$

gives a if $x < a$, b if $x > b$, and x otherwise (Mangel and Clark 1988).

Equation 1 can be understood roughly as follows: If the stimulus at time step t is unfamiliar, then the maximal future expected reproductive success from time $t + 1$ on is given by $F(u(x), t + 1, T)$. With memory length m at time t, this occurs with probability α^m (see Appendix A). Similarly, if the stimulus at time t is familiar and if the organism employs an appropriate response, then the maximal future expected reproductive success from $t + 1$ on is $F(c(x), t + 1, T)$, whereas an inappropriate response at time t would give $F(i(x), t + 1, T)$. With memory length m at time t, the probabilities of appropriate and inappropriate responses to a familiar stimulus are $\omega(1 - \beta^m)$ and $1 - \alpha^m - \omega(1 - \beta^m)$, respectively. Thus, the quantity in braces in Equation 1 corresponds to the future expected reproductive success of an organism with memory length m at time t given that the organism survives the time step (which occurs with probability $1 - \delta$) and remembers optimally from time $t + 1$ onward. To calculate the "maximal" expected future reproductive success at time t (i.e., $F(x, t, T)$), we simply need to find the memory length that maximizes the weighted average in braces—see Appendix A for details.

To begin, we assume that our organism is semelparous—it reproduces at the end of its lifetime (at age T). The expected number of offspring will be some function of the final state of the parent—we call this function $\Phi(x)$. At age T, maximal future reproductive success is equal to the expected number of offspring such that $F(x, T, T) = \Phi(x)$. If we specify $\Phi(x)$, then we can use recursion (1) to obtain $F(x, T - 1, T)$ for all possible values of x. In the process of calculating $F(x, T - 1, T)$, we will also reveal $m^*(x, T - 1)$. This optimal memory (for a given x value) is simply the m value that maximizes the quantity in braces on the right hand side of Equation 1 with $t = T - 1$. Once we have discovered $F(x, T - 1, T)$ for all values of x, we can find both $F(x, T - 2, T)$ and $m^*(x, T - 2)$ for all values of x by applying recursion (1) again (with $t = T - 2$). Then it is a simple matter to generate both

$F(x, T - 3, T)$ and $m^*(x, T - 3)$ using recursion (1) yet again. We can do this over and over again, so that these generated F values become the stepping stones enabling us to walk backward further and further in the organism's lifetime, revealing the optimal memory for each possible age and each possible state along the way. See Table 1 for a summary of the parameters, variables, and functions we have used.

RESULTS

In Figure 1, we give some examples of the optimal memory surface (the m^* function) for different rates of environmental change (ρ) and final fitness functions (Φ). In general, as the stimuli become more reliable (as ρ decreases), optimal memory increases (the surfaces tend to be higher). This is expected to occur because, as ρ decreases, the probability of employing a correct response increases. Thus, there is increasing value to approaching a stimulus as familiar rather than unfamiliar because the remembered response becomes more likely to be correct. This means there is increasing incentive to lengthen the memory window.

For many parameter combinations, we can partition the lifetime into 2 periods: a memory lift at the beginning of the lifetime and a memory plateau in the rest of the lifetime that gently drops or lifts toward the end of the lifetime (see Figure 1). For intermediate values of stimulus change ($0.05 < \rho < 0.25$), if Φ is convex ($d^2\Phi/dx^2 > 0$), the optimal memory window can increase at the end of the lifetime (Figure 1c,d). On the other hand, if Φ is concave ($d^2\Phi/dx^2 < 0$), then the optimal memory window can actually decrease at the end of the lifetime (Figure 1a,b).

Proxy for optimal memory

To understand the effects of stimulus change (ρ), payoff structure (the π's), and the final fitness function (Φ), it helps to consider the following function:

Table 1
Table of parameters, variables, and functions used in the model

| Parameters	Description
X	Number of states for the organism
T	Maximum lifetime for the organism (total number of time steps)
N	Number of stimuli in the environment
ρ	Probability of stimulus change per time step
κ	Cost of living for the organism per time step
δ	Probability of death for the organism per time step
π_u	Payoff to the organism for a response to an unfamiliar stimulus
π_c	Payoff to the organism for a correct response to a familiar stimulus
π_i	Payoff to the organism for an incorrect response to a familiar stimulus
Variables	
x	State of organism (current level of energy reserves)
t	Age of the organism (current number of time steps)
m	Memory "window" (the number of time steps back that are remembered)
Functions	
$u(x)$	State of organism after employing a response to an unfamiliar stimulus, given that its state before responding was x
$c(x)$	State of organism after employing a correct response to familiar stimulus, given that its state before responding was x
$i(x)$	State of organism after employing an incorrect response to familiar stimulus, given that its state before responding was x
$F(x, t, T)$	Maximal expected future reproductive success of an organism in state x at age t
$\Phi(x)$	Fitness of an organism in state x at age T
$H(x)$	The function to which $F(x, t, T)$ converges as t decreases
$B(x, t)$	Potential benefit of memory for an organism in state x at age t
$C(x, t)$	Potential cost of memory for an organism in state x at age t
$m^*(x, t)$	Optimal memory for an organism in state x at age t (an integer)
$m^{**}(x, t)$	Proxy for the optimal memory for an organism in state x at age t (a real number)
$m_H(x)$	Approximate optimal memory (the m^{**} proxy using $H(x)$ in lieu of $F(x, t, T)$)
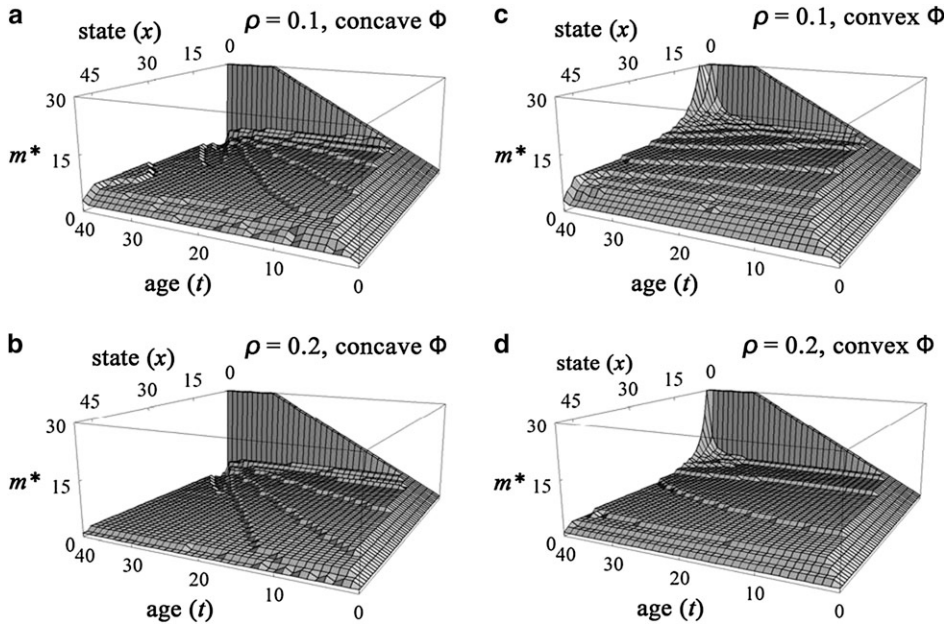
Figure 1
Optimal memory $m^*(x, t)$ with $X = 50$, $N = 10$, $T = 40$, $\delta = 0.1$, $\kappa = 12$, $\pi_i = 5$, $\pi_u = 10$, and $\pi_c = 15$. Each surface corresponds to a different value of the rate of stimulus change (ρ) and either a convex final fitness function $\Phi(x) = 100(x/X)^4$ or a concave final fitness function $\Phi(x) = 100(x/X)^{0.25}$.

$$G(m,x,t) = \left\{ \begin{array}{l} \alpha^m F(u(x), t+1, T) \\ + \omega[1 - \beta^m]F(c(x), t+1, T) \\ + (1 - \omega[1 - \beta^m] - \alpha^m)F(i(x), t+1, T) \end{array} \right\},$$

(6)

with α, β, and ω given above in Equation 2. From Equation 1, the integer m value (with $0 \leq m \leq t$) that gives a maximum for $G(m, x, t)$ is $m^*(x, t)$. One way to find a proxy for m^* is to treat $G(m, x, t)$ as a continuous differentiable function of m and find critical points (by setting $\partial G/\partial m = 0$). A single critical point exists at

$$m^{**}(x,t) = \frac{\ln(1 + v(x,t)) + \ln\frac{\omega\ln\beta}{\ln\alpha}}{\ln\frac{1}{1-\rho}},$$

(7)

with $v(x, t)$ being a benefit-to-cost ratio,

$$v(x,t) = \frac{B(x,t)}{C(x,t)} = \frac{F(c(x), t+1, T) - F(u(x), t+1, T)}{F(u(x), t+1, T) - F(i(x), t+1, T)}.$$

(8)

The potential benefit of using memory is given by the improvement in future expected reproductive success by employing a correct response to a familiar stimulus in lieu of a response to an unfamiliar stimulus—this is precisely $B(x, t)$ from Equation 8. The potential cost of using memory is given by the loss in future expected reproductive success by employing an incorrect response to a familiar stimulus in lieu of learning over a single time step in response to an unfamiliar stimulus—this is precisely $C(x, t)$ from Equation 8. Thus, without attention to the probabilities of correct and incorrect responses, v gives a benefit-to-cost ratio of memory use.

The quantity m^{**} in Equation 7 is a proxy for the actual optimal memory window (m^*, which must be an integer). If m^{**} is between 0 and t, then the integer below or above m^{**} gives m^* (m^* is the integer which yields a larger value for Equation 6). If m^{**} is negative, then optimal memory is zero, and if m^{**} is greater than t, then optimal memory is t (see Supplementary Appendix A).

This proxy for optimal memory depends not only on the benefit-to-cost ratio (v) but also on the probabilities that memory

brings benefits versus costs. These probabilities depend on the number of stimuli, N, and the rate of stimulus change, ρ (which is why there are α, β, and ω terms in Equation 7—see Equation 2). In Figure 2, we show how the probabilities of incorrect familiar responses (orange bars), unfamiliar responses (green bars), and correct familiar responses (blue bars) change with the size of the memory window for different values of ρ. From Equation 1, as m grows larger, the probability of a correct response approaches ω, the probability of an incorrect response approaches $1 - \omega$, and the probability of an unfamiliar response goes to zero. As ρ increases, the probabilities of incorrect responses at any memory level also increase (the fraction of orange increases), which should favor lower memory values. Indeed, we show in Supplementary Appendix B that when our optimal memory proxy, m^{**}, is greater than 0, increasing the rate of stimulus change (i.e., decreasing reliability) will promote shorter memory windows, that is, $\partial m^{**}/\partial\rho < 0$. This result has been derived by treating m in function G as continuous; however, when m is constrained to integer values, the same general pattern emerges: stimulus changeability promotes shorter memory windows (see Figure 1).

Also, m^{**} is an increasing function of v (i.e., $\partial m^{**}/\partial v > 0$). In order to explore the effects of v, we need to know something about $F(x, t, T)$. Although the precise analytical form of $F(x, t, T)$ is difficult to derive, it will generally be a nondecreasing function of x, given that we assume $\Phi(x)$ is a nondecreasing function of x. For illustrational purposes in the next section, we will pretend F assumes simple functional forms (e.g., a linear function); however, this pretence will not affect most of the conclusions for more complicated nondecreasing forms of F.

Graphical approach

Let us focus on an organism in state x' at age t (the black circles in Figure 3). If it employs an incorrect response to a familiar stimulus, its state changes to $i(x')$ (the orange solid arrows in Figure 3). If it employs a response to an unfamiliar stimulus, its state changes to $u(x')$ (the green dotted lines and arrows). And if it employs a correct response to a familiar stimulus, its state changes to $c(x')$ (the blue dashed lines and arrows). We always have $i(x) \leq u(x) \leq c(x)$ because we assume
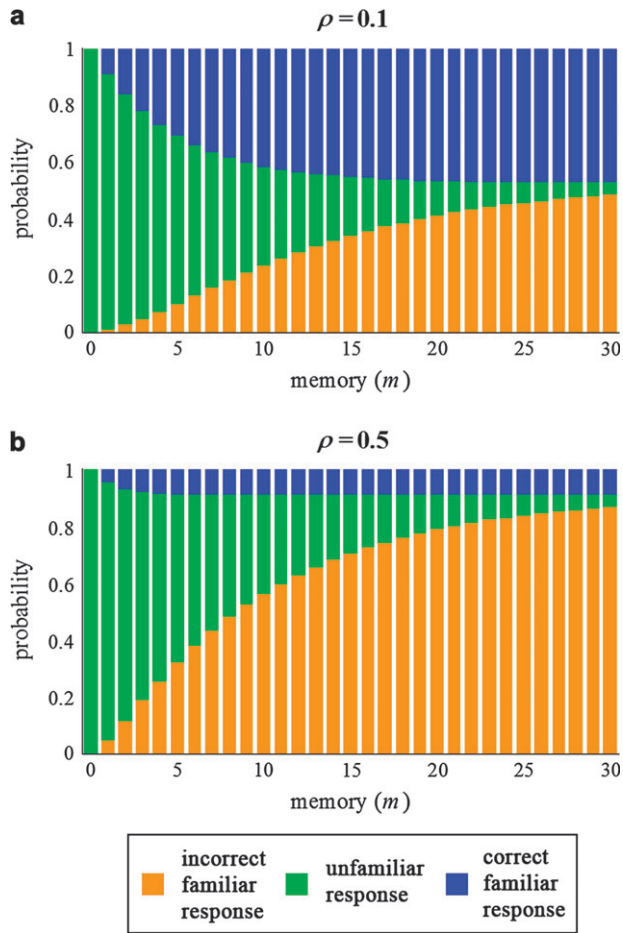
Figure 2
The probability of a correct response to a familiar stimulus (blue bars), the probability of a response to an unfamiliar stimulus (green bars), and the probability of an incorrect response to a familiar stimulus (orange bars) as a function of memory size (with $N = 10$). Each chart corresponds to a different value of the rate of stimulus change, ρ. These probabilities are $\omega(1 - \beta^m)$, α^m, $1 - \alpha^m - \omega(1 - \beta^m)$, respectively.

$\pi_i < \pi_u < \pi_c$. In each case, the maximal future expected reproductive success is given by the value of $F(x, t + 1, T)$ at $x = i(x')$, $x = u(x')$, or $x = c(x')$. From the positions of $c(x')$, $u(x')$, and $i(x')$, we can visualize the benefit, $B(x, t)$, and cost, $C(x, t)$, of memory use and thus isolate factors that will affect ν.

We see that for constant values of correct and incorrect payoffs, π_c and π_i (i.e., $c(x')$ and $i(x')$ remain constant), as the value of the payoff for an unfamiliar stimulus, π_u, decreases (i.e., as $u(x')$ decreases), ν increases and longer memory windows are favored (compare Figure 3a,b). The quantity $\pi_u - \pi_i$ generally scales with the potential cost of memory, $C(x, t)$. The potential benefit of memory scales with $\pi_c - \pi_u$. It stands to reason that as the potential cost of memory decreases and the potential benefit of memory increases (π_u drops while π_c and π_i both stay constant), the optimal memory window should be longer.

We can also start to understand some of the effects of the state of the organism on memory. Given that we assume that the cost of living is high (i.e., $\kappa > \pi_u$), if the state of the organism is very low, then both $i(x')$ and $u(x')$ will be zero (remember these functions "chop" x into the range of 0 to X). Again, the cost of living is the energy reserves removed from the organism due to the metabolic demands of living. This can be seen in Figure 3c, where the solid and dotted
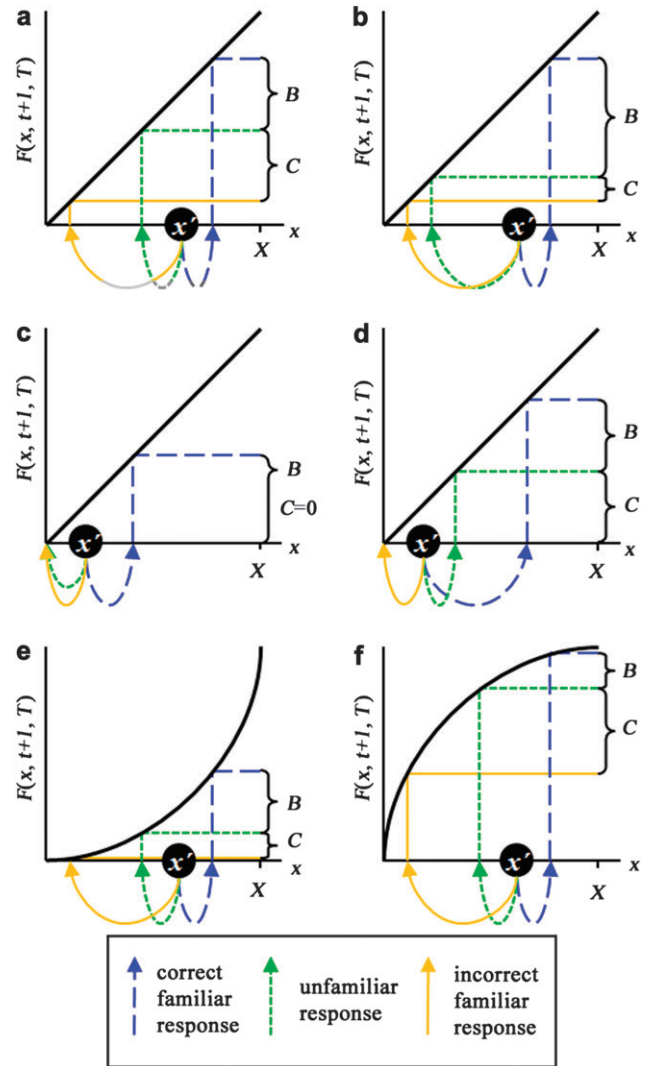

Figure 3
Schematics to show the costs and benefits of memory. (a) We start by assuming that $F(x, t + 1, T)$ is a linear function (although generally it is not). We assume that the state of our organism at age t is x'. A solid orange arrow points to $i(x')$, the state acquired through an incorrect response to a familiar stimulus. A dotted green arrow points to $u(x')$, the state acquired through a response to an unfamiliar stimulus. A dashed blue line and arrow points to $c(x')$, the state acquired through a correct response to a familiar stimulus. The potential benefit $B(x', t)$ of being familiar with a stimulus is $F(c(x'), t + 1, T) - F(u(x'), t + 1, T)$ and potential cost $C(x', t)$ of being familiar with a stimulus is $F(u(x'), t + 1, T) - F(i(x'), t + 1, T)$. (b) If π_u decreases, then $u(x')$ can decrease and C decreases. (c) If the state of the organism is very low and the cost of living is very high ($\kappa > \pi_u$), then $i(x') = u(x') = 0$ while $c(x') > 0$. Consequently, $C = 0$ and $B > 0$, that is, there are no costs to memory, whereas there are benefits—thus, the memory window should be maximally long. (d) If the cost of living is low ($\kappa < \pi_u$, see Supplementary Appendix C), then $C > 0$ and $B > 0$ and maximal memory may not be favored. (e) If $F(x, t + 1, T)$ is convex, the ratio of B to C increases (relative to the linear F function in (a)), and thus an increased memory window is favored. (f) If $F(x, t + 1, T)$ is concave, the ratio of B to C decreases (relative to the linear function in (a)), and thus a decreased memory window is favored.

arrows point to the origin. Because $i(x') = u(x') = 0$, this means that $C(x', t) = 0$ and ν is infinite. As $\nu \to \infty$, we know that $m^{**} \to \infty$. Thus, when cost of living is high, organisms in a very low state should possess maximal memory. Again the

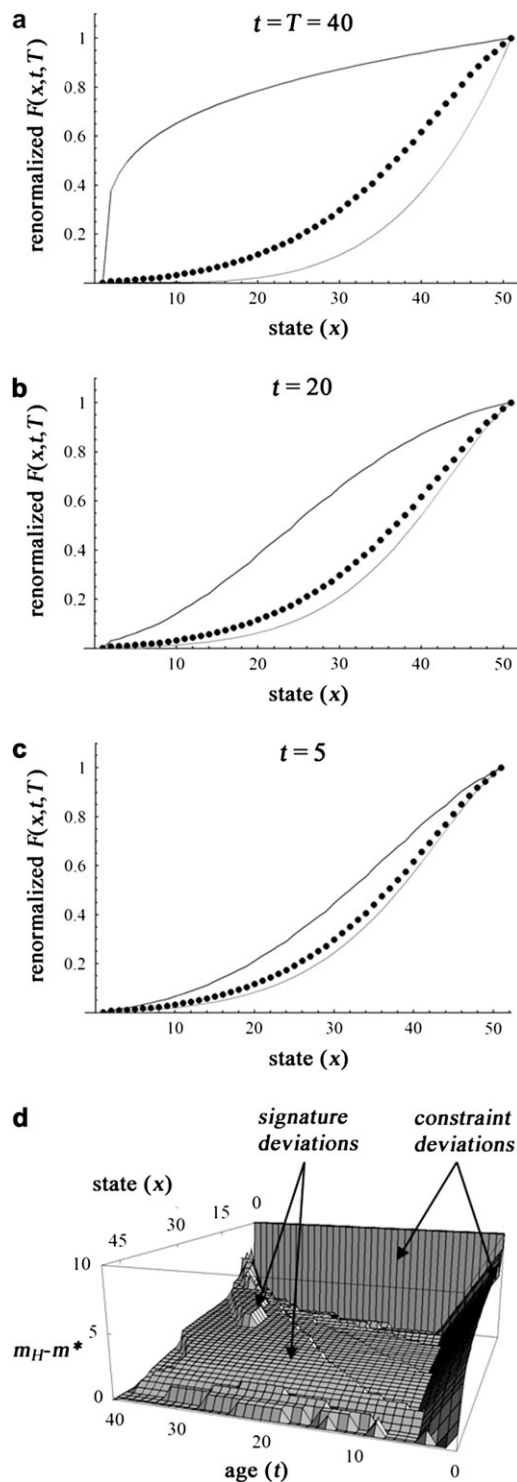idea is that employing a correct response to a familiar stimulus is the only way to survive; thus, there is a premium placed on a maximal memory window. If we had assumed that the costs of living are lower (i.e., $\kappa < \pi_u$), then $u(x') > 0$ for an organism in a very low state and optimal memory is no longer necessarily maximal (see Figure 3d; in Supplementary Appendix C, we explore the case when $\kappa < \pi_u$).

The benefit-to-cost ratio of memory, ν, also depends on the shape of the expected future reproductive value function, F. For instance, ν increases as F is made more convex (compare Figure 3e with a). Thus, optimal memory windows are longer for convex F functions and shorter for concave F functions (compare Figure 3f with a). This prompts the question: what is the general shape of the F function at any arbitrary age of our organism? For one age, the answer to this question is clear. At the end of the organism's lifetime, $F(x, T, T) = \Phi(x)$ and we specify this final fitness function. In Figure 1, we see that a convex Φ leads to a longer optimal memory window at the end of the lifetime than a concave Φ (see Supplementary Appendix D).

Convergence

As mentioned, an analytical solution of the general form of $F(x, t, T)$ is difficult to derive. However, we do observe a convergence property in our model that is common to many dynamic programming routines (Mangel and Clark 1988; Houston and McNamara 1999). As we step backward in time through repeated application of Equation 1, $F(x, t, T)$ assumes a form (once renormalized) that is independent of the final fitness function, $\Phi(x)$. If we do not constrain m to remain below t (mathematically possible, but biologically unjustified), this function is also independent of time (this is a form of strong convergence, see Houston and McNamara 1999). The convergence is shown in Figure 4 for the examples from Figure 1a,c. If we assume that F does strongly converge to some function, call it $H(x)$, then the proxy for optimal memory $m^{**}(x, t)$ is approximated by replacing $F(x, t, T)$ in ν with $H(x)$ in Equation 7. Let us call this approximate optimal memory $m_H(x)$. We show an example of $H(x)$ in black dots in Figure 4a–c. For a given x value, $m^*(x, t) \approx m_H(x)$ for several different t values.

However, there are a few ages (t values) where the $m_H(x)$ approximation fails. First, for any state x, if $t < m_H(x)$, we know the memory must be smaller than the approximation because the organism cannot remember stimuli from before it was born. Generally, for t values below $m_H(x)$, we find $m^*(x, t) = t$, that is, the optimal memory window is the maximum in the allowable range (incidentally, the constraints on m, which were ignored to compute $H(x)$, force $F(x, t, T)$ to deviate from $H(x)$ for the young ages $t < m_H(x)$). Second, for ages close to the end of the lifetime, $t \approx T$, there is a strong signature of $\Phi(x)$ on $F(x, t, T)$ (i.e., backward convergence to $H(x)$ has not proceeded very far). If $\Phi(x)$ is more convex than $H(x)$, then the optimal memory window bends up at the end of the

Figure 4
(a) The final fitness function $\Phi(x) = 100(x/X)^4$ is given in gray and $\Phi(x) = 100(x/X)^{0.25}$ is given in black. These functions are normalized to range from 0 to 1. (b) The corresponding renormalized $F(x, t, T)$ functions at time $t = 20$, generated through backward iteration of Equation 1 (with $X = 50$, $N = 10$, $T = 40$, $\delta = 0.1$, $\pi_i = 5$, $\pi_u = 10$, $\pi_c = 15$, $\kappa = 12$, and $\rho = 0.1$) starting with $F(x, T, T) = 100(x/X)^4$ (in gray) and $F(x, T, T) = 100(x/X)^{0.25}$ (in black). (c) The corresponding renormalized $F(x, t, T)$ functions at $t = 5$. Both functions are approaching an equilibrium function $H(x)$ given in black dots in (a–c). (d) Using $H(x)$ in place of $F(x, t + 1, T)$, we calculate $m_H(x, t)$ by Equation 7 for all x and t values. Here we show the difference between m_H and m^* to show the appropriateness

of the m_H approximation. We see 2 "types" of deviations between m_H and m^*. The constraint deviations occur where the m_H approximation is outside of the allowed range for memory (e.g., $m_H(x, t) > t$) and the signature deviations occur when m_H is near attainable memory values, but other memory values are optimal (these deviations are due to the fact that $H(x)$ is an inappropriate approximation for $F(x, t, T)$). In the case shown, the signature deviations ($m_H - m^*$) are positive, suggesting that there is optimal decline in memory for large values of t (note that we are dealing with the concave final fitness function shown in black in a).

lifetime. However, if $\Phi(x)$ is more concave than $H(x)$, then the optimal memory window bends down at the end of the lifetime. In such a case, it would be optimal to reduce long-term recall at the end of the lifetime, that is, memory loss would be optimal.

We can now start to understand the age-related phases of optimal memory that we saw in Figure 1. The memory lift at the beginning of the lifetime occurs because the young age of the organism limits its recall. We call this a "constraint deviation" in Figure 4d because the young organism is constrained by its age to remember less than $m_H(x)$. The other constraint deviation is for organisms in low state where $m_H(x)$ is predicted to be infinite, but the organism must have finite memory. The memory plateau begins at ages larger than $m_H(x)$ and gently slopes toward the end of the lifetime. The deviations here we call "signature deviations" because the signature of the final fitness function is echoed in the optimal memory at the end of the organism's lifetime, where Φ functions that are convex relative to $H(x)$ lift final optimal memory and relatively concave Φ functions (e.g., like that used to produce Figure 4d) depress final optimal memory.

DISCUSSION

Using a dynamic programming approach, we have suggested that optimal retrieval of memories should not be expected to be constant but rather should change with the age and state of the organism. Thus, this model proposes that specific changes in memory (e.g., loss of memory retrieval toward the end of the lifetime) could be adaptive. Of course, these findings do not deny that there may be other nonadaptive factors or constraints that affect memory retrieval, such as the general deterioration of memory for older individuals due to inevitable breakdown of neural function. Rather, this model simply states that it is not necessary to invoke such constraints and that an organism that maximizes its future reproductive output is expected to alter its retrieval of memories under certain circumstances.

We should note that, even within the context of our model, there are changes in memory retrieval that result from constraints. For instance, an organism is forced to have shorter memory windows when it is younger because it cannot have windows that exceed its age. Mathematically, we could iterate recursion (1) without constraining the memory window, m, to be less than the current age of the organism, t. If m can assume any positive value, we have a situation that an organism can "remember" events before it was born. This exercise, although biologically unrealistic, reveals that our convergence approximation, m_H, is a good approximation of the optimal m value all the way to the youngest age, thus demonstrating that constraints keep the memory window small early in the lifetime.

However, this same mathematical exercise does not affect results from later in life. Thus, it is not constraints on m that lead to memory decline or increase toward the end of the lifetime. Why is optimal memory changing here? Within our model, memory generates a variance in payoffs. This means that the concavity of the maximal expected future reproductive success function, F, will be important. Specifically, the spread in payoffs are translated positively for convex F and negatively for concave F. Of course, F changes with age. Specifically, for the oldest age, F is given by Φ, the final fitness function, and as age decreases, F converges to another function, which we have called H. If Φ is more convex than H, then memory tends to increase toward the end of the lifetime. If Φ is more concave than H, then memory tends to decrease toward the end of the lifetime. When H does not have much curvature (e.g., Figure 4), then strongly convex or concave

final fitness functions will tend to leave "signatures" on the optimal memory surface (gradual increase or gradual decline, respectively). The question remains whether a convex or concave final fitness function is more biologically realistic. If fitness returns consistently diminish (as opposed to accelerate) with energy reserves, then a concave function would be apt. In such a case, our model predicts that optimal memory should decline with age, an effect due to the signature of the fitness function rather than constraints on memory.

Many theoretical treatments on the evolutionary advantages of learning center on environmental variability (Arnold 1978; Johnston 1982; McNamara and Houston 1987; Stephens 1987, 1991; Mangel 1990; Cohen 1991; Bergman and Feldman 1995; Feldman et al. 1996; Dukas 1998; Shettleworth 1998; Kerr and Feldman 2003). Consistent with other models, our model shows that as stimuli in the environment become less reliable, optimal recall (at any age or state) tends to decrease. However, the rate of stimulus change, ρ, will also have a strong effect on the shape of H, the function to which the maximal expected future reproductive success function converges. As the concavity of H changes, this will affect the differences in concavity between Φ and H and thus will change the potential for optimal memory decline or increase (see above).

When the organism is in a low state (and the cost of living is high, which we assume in our model), maximal memory is optimal. Consistent with this prediction, Pravosudov and Clayton (2001) found that mountain chickadees maintained on a limited and unpredictable food supply made fewer errors on a cache recovery task than well-fed birds (see also Friedrich et al. 2004; Orsini et al. 2004). Of course, the motivation of an organism with low energy levels and high costs of living might be very different than that of an organism with high energy levels and low costs of living, and cognitive performance may vary with motivation.

Although this model provides some novel insights, there are a number of potential extensions that could be explored. First, reproduction in the model occurs only at the end of the lifetime (a form of semelparity). It is not difficult to incorporate reproduction throughout the lifetime (Mangel and Clark 1988; Houston and McNamara 1999). We explore an iteroparous extension to our model in Supplementary Appendix E, where we show that optimal memory can still decay with age despite repeated individual bouts of reproduction. Thus, adaptive memory decay does not depend on an assumption of semelparity. Second, we assumed that the cost of living, κ, was constant. However, it has been proposed that memory itself may be explicitly costly (Dukas 1999; Mery and Kawecki 2005; Kuhl et al. 2007), and memory performance is sensitive to imposed costs (Bednekoff and Balda 1997; Laughlin and Mendl 2004). We could easily include cost by writing κ as an increasing function of memory ($\kappa(m)$). We consider this model extension in Supplementary Appendix F. Although explicit costs to memory depress optimal values, we see that optimal memory can still decay with age. With explicit memory costs, we can also reorder the payoffs, such that using memory is always more valuable than sampling from scratch: $\pi_u < \pi_i < \pi_c$. Such an ordering could apply to the case where stimulus change leads to only a slight decrease in payoff for a remembered behavior. Again, we see that optimal memory can still decrease with age in such a case (Supplementary Appendix F). We have found that our basic results are robust to alterations in several of our model assumptions (Supplementary Appendices E and F).

Our model offers a new perspective on the evolutionary reasons for deterioration of memory with age. The classic theories of senescence (Medawar 1952; Williams 1957) suggest that deterioration with age occurs either due to selection for genes that are beneficial early in life but detrimental later in life

Table 2
Broad predictions from the model

Parameter/variable	Condition	Optimal memory length
Probability of stimulus change (ρ)	Stimuli become less reliable	Decreases
Cost of living (κ)	Cost of living increases	Increases
State of organism (x)	Poor state, high cost of living	Long
	Good state, low cost of living	Short
Age (t)	Relatively convex final fitness function	Increases at the end of life
	Relatively concave final fitness function	Decreases at the end of life

(antagonistic pleiotropy) or due to the inefficacy of selection to weed out deleterious mutations late in life when reproductive output decreases (mutation accumulation). These evolutionary explanations generally posit that, all else being equal, decay at any age is selected against (and thus senescence is either a side effect or a selectively elusive phenomenon). Some of the classic explanations of memory loss (e.g., synaptic decay) fit well within this classic theoretical framework. In contrast, our model claims that memory loss can in and of itself be adaptive. We expect this to be the case when condition is poor, the individual is near the end of its lifetime, and fitness returns consistently diminish with energy reserves.

Furthermore, our model makes several specific predictions about changes in memory (Table 2). First, and in agreement with other models, memory length should decrease as the environment becomes less reliable (McNamara and Houston 1987; Mangel 1990; Kerr and Feldman 2003). Second, memory length should roughly increase with the cost of living (see Supplementary Appendix C). Third, cost of living should interact with state of the organism. If the cost of living is high, organisms in a very poor state should have greater memory length, whereas if the cost of living is low, organisms in a very good state should have short to no memory length. Fourth (and unlike general theories of senescence), our model does not predict inevitable decline of memory length. The behavior of memory length with age depends on the relationship between the state of the organism and its final fitness. However, if fitness shows diminishing returns with state, memory should decline gently with age for intermediate levels of stimulus reliability.

The validity of this model can be addressed empirically. For organisms at reproductive age, the relationship between energy (or fat) reserves and number of offspring is an empirical issue. The predicted changes in memory could be checked (perhaps within a comparative framework) against the observed concavity of this relationship. Individuals with high metabolic demands of living tend to demonstrate superior memory capability while in poor states (e.g., Pravosudov and Clayton 2001; Friedrich et al. 2004; Orsini et al. 2004). It would be interesting to repeat such experiments with organisms that possess a low cost of living (where our model predicts that the effect will be less dramatic). Given the rich history of empirical work on memory, models such as this one will be important in both reconciling seemingly disparate observations and suggesting new avenues of research in the behavioral ecology of memory. The essential point underscored by this model is that learning is a dynamic process and critical components of that process, such as memory length, may change in optimal ways within the lifetime of the learner.

SUPPLEMENTARY MATERIAL

Supplementary material can be found at http://www.beheco. oxfordjournals.org/.

APPENDIX

A: Derivation of the dynamic programming recursion

In order to derive recursion (1), we list the possible events that can occur at time step t to an organism in state x. Let **U** be the event that our organism is unfamiliar with the stimulus it experiences. Let **C** be the event that our organism is familiar with the stimulus and employs the correct behavior (i.e., the correct behavior to the stimulus has not changed since its most recent recording in the organism's memory). Let **I** be the event that our organism is familiar and employs an incorrect behavior (i.e., the correct behavior to the stimulus has changed since its most recent recording in the organism's memory). We let the dummy event variable, **A**, stand for one of the stimulus encounter events, that is, $\mathbf{A} \in \{\mathbf{U}, \mathbf{C}, \mathbf{I}\}$. Let $\mathbf{X}_x$ be the event that the organism's state is x, $\mathbf{T}_t$ be the event that an organisms age is t, and $\mathbf{M}_m$ be the event that an organism's memory window is m.

Consider that our organism in state x at age t has memory size m, where this memory window may not be optimal. After this time step, we assume that the organism adjusts its memory window optimally (i.e., we assume $m(x, \tau) = m^*(x, \tau)$ for all $\tau \geq t + 1$). We call this organism the "nearly optimal" learner. Let the expected future reproductive success of the nearly optimal learner be given by $F_m(x, t, T)$. In order to calculate F_m, we let $EF[\mathbf{A}|\mathbf{X}_x \cap \mathbf{T}_t]$ be the expected future reproductive success for the nearly optimal learner if event **A** occurs (given that the organism's state is x and its age is t). We let $\Pr\{\mathbf{A}|\mathbf{M}_m\}$ be the probability of event **A**, given our organism has a memory window of size m. By the law of total expectation, we must have

$$F_m(x, t, T) = \sum_{\mathbf{A} \in \{\mathbf{U}, \mathbf{C}, \mathbf{I}\}} EF[\mathbf{A}|\mathbf{X}_x \cap \mathbf{T}_t] \Pr\{\mathbf{A}|\mathbf{M}_m\}, \qquad \text{(A1)}$$

Our nearly optimal learner becomes a "completely optimal" learner if it picks the memory window size m (where $0 \leq m \leq t$) that maximizes $F_m(x, t, T)$ (i.e., if it chooses $m = m^*(x, t)$). Thus, we can express the maximal expected future reproductive success of our completely optimal learner in state x at age t as

$$F(x, t, T) = \max_{0 \leq m \leq t} F_m(x, t, T). \qquad \text{(A2)}$$

Now we simply need to derive the expectations and the probabilities in Equation A1. Because we assume that the nearly

optimal learner behaves optimally from time $t + 1$ onward, the expectations are simply

$$EF[\mathbf{U}|\mathbf{X}_x \cap \mathbf{T}_t] = (1 - \delta)F(u(x), t+1, T), \quad (A3)$$

$$EF[\mathbf{C}|\mathbf{X}_x \cap \mathbf{T}_t] = (1 - \delta)F(c(x), t+1, T), \quad (A4)$$

$$EF[\mathbf{I}|\mathbf{X}_x \cap \mathbf{T}_t] = (1 - \delta)F(i(x), t+1, T). \quad (A5)$$

For example, if the stimulus is unfamiliar, the organism receives a payoff of π_u and pays a metabolic cost of κ (with the constraint that the next state is between 0 and X). Thus, if the organism survives from age t to age $t + 1$ (which occurs with probability $1 - \delta$), the organism changes from state x at age t to state $u(x) = \text{chop}(x + \pi_u - \kappa; 0, X)$ at age $t + 1$. Because the organism is assumed to remember optimally from $t + 1$ onward, the expected future reproductive success at $t + 1$ is the maximal expected future reproductive success, which is given by $F(u(x), t + 1, T)$. Thus, given that a dead organism has zero future reproductive success (and given that death occurs with probability δ), the expected reproductive success is $EF[\mathbf{U}|\mathbf{X}_x \cap \mathbf{T}_t] = (1-\delta)F(u(x), t+1, T) + \delta(0) = (1-\delta)F(u(x), t+1, T)$. The other EF values are derived similarly.

Now we turn to the probabilities in Equation A1. We assume that every time step, a stimulus is chosen randomly from the N stimuli, so the probability that the currently experienced stimulus does not occur in any one of the last m time steps must be $(N - 1)/N$. Because we assume that each stimulus occurs independently of previous stimuli, the probability that the current stimulus does not occur in all the previous m time steps is

$$\Pr\{\mathbf{U}|\mathbf{M}_m\} = \left(\frac{N-1}{N}\right)^m. \quad (A6)$$

Equation A6 gives the probability that the organism does not remember the current stimulus (i.e., event $\mathbf{U}$ occurs) given that the organism has a memory window of size m.

We know that without any memory, the stimulus cannot be familiar and therefore $\Pr\{\mathbf{C}|\mathbf{M}_0\} = \Pr\{\mathbf{I}|\mathbf{M}_0\} = 0$. To calculate $\Pr\{\mathbf{C}|\mathbf{M}_m\}$ and $\Pr\{\mathbf{I}|\mathbf{M}_m\}$ when $m > 0$, it helps to condition on other events. Let $\mathbf{F}$ be the event that the stimulus is familiar. because $\mathbf{F}$ is the complement of $\mathbf{U}$, we must have

$$\Pr\{\mathbf{F}|\mathbf{M}_m\} = 1 - \left(\frac{N-1}{N}\right)^m. \quad (A7)$$

Let $\mathbf{S}_r$ be the event that the stimulus was most recently recorded in memory r time steps before the current time step (thus, $\mathbf{S}_r$ is only defined for $r \leq m$). Given that an organism is dealing with a familiar stimulus and has a memory size of m, the probability that the most recently recorded encounter with the current stimulus happened r time steps back is

$$\Pr\{\mathbf{S}_r|\mathbf{F} \cap \mathbf{M}_m\} = \frac{N^{m-r}(N-1)^{r-1}}{N^m - (N-1)^m}. \quad (A8)$$

To derive Equation A8, let us consider the memory window of size m as being constructed of m "slots," ordered from 1 (the most recent memory) to m (the most distant memory). There are a total of $N^m - (N-1)^m$ sequences of stimuli in memory where the current stimulus occurs in at least one of the m slots. How many of those sequences have the current stimulus occurring most recently in slot r? For this to occur, we must have any stimulus except for the current stimulus in each of the slots from slot 1 to slot $r - 1$ (a total of $(N-1)^{r-1}$ combinations for these slots), whereas we can have any stim-

ulus in each of the slots from slot $r + 1$ to slot m (a total of N^{m-r} combinations for these slots). Note that we must have the current stimulus in slot r (a single "combination" for this slot). Thus, the total number of combinations for the current familiar stimulus to be located most recently in the rth slot is $(N - 1)^{r-1}N^{m-r}$ and Equation A8 follows. Given that our organism with memory length m experienced a familiar stimulus most recently r time steps back, the probabilities that the behavior employed is correct is

$$\Pr\{\mathbf{C}|\mathbf{S}_r \cap \mathbf{F} \cap \mathbf{M}_m\} = (1 - \rho)^r, \quad (A9)$$

because the probability of stimulus change every time step is ρ, and change occurs independently each time step. The conditional probability giving an incorrect behavior is calculated easily by noting that $\mathbf{I}|\mathbf{S}_r \cap \mathbf{F} \cap \mathbf{M}_m$ is the complement of $\mathbf{C}|\mathbf{S}_r \cap \mathbf{F} \cap \mathbf{M}_m$ and therefore

$$\Pr\{\mathbf{I}|\mathbf{S}_r \cap \mathbf{F} \cap \mathbf{M}_m\} = 1 - (1 - \rho)^r. \quad (A10)$$

Putting everything together, we have

$$\Pr\{\mathbf{C}|\mathbf{M}_m\} = \sum_{r=1}^{m} \Pr\{\mathbf{C}|\mathbf{S}_r \cap \mathbf{F} \cap \mathbf{M}_m\}\Pr\{\mathbf{S}_r|\mathbf{F} \cap \mathbf{M}_m\}\Pr\{\mathbf{F}|\mathbf{M}_m\}, \quad (A11)$$

$$\Pr\{\mathbf{I}|\mathbf{M}_m\} = \sum_{r=1}^{m} \Pr\{\mathbf{I}|\mathbf{S}_r \cap \mathbf{F} \cap \mathbf{M}_m\}\Pr\{\mathbf{S}_r|\mathbf{F} \cap \mathbf{M}_m\}\Pr\{\mathbf{F}|\mathbf{M}_m\}. \quad (A12)$$

Substituting Equations A7–A9 into Equation A11 and substituting Equations A7, A8, and A10 into Equation A12 and then simplifying gives

$$\Pr\{\mathbf{C}|\mathbf{M}_m\} = \frac{1-\rho}{1 + (N-1)\rho}\left[1 - \left(\frac{(N-1)(1-\rho)}{N}\right)^m\right], \quad (A13)$$

$$\Pr\{\mathbf{I}|\mathbf{M}_m\} = 1 - \left(\frac{(N-1)}{N}\right)^m - \frac{1-\rho}{1 + (N-1)\rho}\left[1 - \left(\frac{(N-1)(1-\rho)}{N}\right)^m\right]. \quad (A14)$$

Now, plugging Equations A3–A6, A13, and A14 into Equation A1 and then plugging Equation A1 into Equation A2 yields Equation 1, our backward recursion.

REFERENCES

Anderson MC, Ochsner KN, Kuhl B, Cooper J, Robertson E, Gabrieli SW, Glover GH, Gabrieli JDE. 2004. Neural systems underlying the suppression of unwanted memories. Science. 303:232–235.

Arnold SJ. 1978. The evolution of a special class of modifiable behaviors in relation to environmental pattern. Am Nat. 112:415–427.

Bednekoff PA, Balda RP. 1997. Clark's nutcracker spatial memory: many errors might not be due to forgetting. Animal Behaviour. 54: 691–698.

Bergman A, Feldman MW. 1995. On the evolution of learning: representation of a stochastic environment. Theor Popul Biol. 48:251–276.

Bouton ME. 1994. Conditioning, remembering, and forgetting. J Exp Psychol Anim Behav Process. 20:219–231.

Bouton ME, Moody EW. 2004. Memory processes in classical conditioning. Neurosci Biobehav Rev. 28:663–674.

Clark CW. 1993. Dynamic models of behavior: an extension of life history theory. Trends Ecol Evol. 8:205–209.

Cohen D. 1991. The equilibrium distribution of optimal search and sampling effort of foraging animals in patchy environments. In:

Yoshimura J, Clark CW, editors. Adaptation in stochastic environments. Berlin (Germany): Springer-Verlag. p. 173–191.

Cohen G. 1996. Memory and learning in normal aging. In: Woods RT, editor. Handbook of the clinical psychology of ageing. London: Wiley. p. 43–58.

Craik FIM, Jennings JM. 1992. Human memory. In: Craik FIM, Salthouse TA, editors. The handbook of aging and cognition. Hillsdate (NJ): Erlbaum. p. 51–110.

Davis HP, Small SA, Stern Y, Mayeux R, Feldstein SN, Keller FR. 2003. Acquisition, recall, and forgetting of verbal information in long–term memory by young, middle-aged, and elderly individuals. Cortex. 39:1063–1091.

Davis HP, Trussell LH, Klebe KJ. 2001. A ten-year longitudinal examination of repetition priming, incidental recall, free recall, and recognition in young and elderly. Brain Cogn. 46:99–104.

Dukas R. 1998. Evolutionary ecology of learning. In: Dukas R, editor. Cognitive ecology: the evolutionary ecology if information processing and decision making. Chicago: University of Chicago Press. p. 129–164.

Dukas R. 1999. Costs of memory: ideas and predictions. J Theor Biol. 197:41–50.

Feldman MW, Aoki K, Kumm J. 1996. Individual versus social learning: evolutionary analysis in a fluctuating environment. Anthropol Sci. 104:209–231.

Friedrich A, Thomas U, Muller U. 2004. Learning at different satiation levels reveals parallel functions for the cAMP-protein kinase A cascade in formation of long-term memory. J Neurosci. 24:4460–4468.

Guo A, Li L, Xia SZ, Feng CH, Wolf R, Heisenberg M. 1996. Conditioned visual flight orientation in Drosophila: dependence on age, practice, and diet. Learn Mem. 3:49–59.

Hirvonen H, Ranta E, Rita H, Peuhkuri N. 1999. Significance of memory properties in prey choice decisions. Ecol Modell. 115:177–189.

Houston AI, McNamara JM. 1999. Models of adaptive behavior: an approach based on state. Cambridge (UK): Cambridge University Press.

Johnston TD. 1982. Selective costs and benefits in the evolution of learning. Adv Study Behav. 12:65–106.

Kerr B, Feldman MW. 2003. Carving the cognitive niche: optimal learning strategies in homogeneous and heterogeneous environments. J Theor Biol. 220:169–188.

Kraemer PJ, Golding JM. 1997. Adaptive forgetting in animals. Psychon Bull Rev. 4:480–491.

Kuhl BA, Dudukovic NM, Kahn I, Wagner AD. 2007. Decreased demands on cognitive control reveal the neural processing benefits of forgetting. Nat Neurosci. 10:908–914.

Laughlin K, Mendl M. 2004. Costs of acquiring and forgetting information affect spatial memory and its susceptibility to interference. Animal Behaviour. 68:97–103.

Lewis AC. 1986. Memory constraints and flower choice in Pieris-Rapae. Science. 232:863–865.

Loftus EF, Loftus GR. 1980. On the permanence of stored information in the brain. Am Psychol. 35:409–420.

Lupien SJ, de Leon M, De Santi S, Convit A, Tarshish C, Nair NPV, Thakur M, McEwan BS, Hauger RL, Meaney MJ. 1998. Cortisol levels during human aging predict hippocampal atrophy and memory deficits. Nat Neurosci. 1:69–73.

Mangel M. 1990. Dynamic information in uncertain and changing worlds. J Theor Biol. 146:317–332.

Mangel M, Clark CW. 1988. Dynamic modeling in behavioral ecology. Princeton (NJ): Princeton University Press.

McEwen BS, Sapolsky RM. 1995. Stress and cognitive function. Curr Opin Neurobiol. 5:205–216.

McGaugh JL, Roozendaal B. 2002. Role of adrenal stress hormones in forming lasting memories in the brain. Curr Opin Neurobiol. 12:205–210.

McNamara JM, Houston AI. 1987. Memory and the efficient use of information. J Theor Biol. 125:385–395.

Medawar PB. 1952. An unsolved problem of biology. London: H.K. Lewis.

Mery F, Kawecki TJ. 2004. An operating cost of learning in Drosophila melanogaster. Anim Behav. 68:589–598.

Mery F, Kawecki TJ. 2005. A cost of long-term memory in Drosophila. Science. 308:1148.

Neissen U. 2004. Memory development: new questions and old. Dev Rev. 24:154–158.

Orsini C, Buchini F, Conversi D, Cabib S. 2004. Selective improvement of strain-dependent performances of cognitive tasks by food restriction. Neurobiol Learn Mem. 81:96–99.

Pravosudov VV, Clayton NS. 2001. Effects of demanding foraging conditions on cache retrieval accuracy in food-caching mountain chickadees (Poecile gambeli). Proc R Soc Lond B Biol Sci. 268: 363–368.

Punzo F, Chavez S. 2003. Effect of aging on spatial learning and running speed in the shrew (Cyptotis parva). J Mammal. 84:1112–1120.

Salthouse TA. 1996. The processing-speed theory of adult age differences in verbal tasks. Psychol Rev. 103:403–428.

Schacter DL. 1999. The seven sins of memory: insights from psychology and cognitive neuroscience. Am Psychol. 54:182–203.

Schlaepfer MA, Runge MC, Sherman PW. 2002. Ecological and evolutionary traps. Trends Ecol Evol. 17:474–480.

Seeman TE, McEwan BS, Singer BS, Albert MS, Rowe JW. 1997. Increase in urinary cortisol excretion and memory declines: MacArthur studies of successful aging. J Clin Endocrinol Metab. 82: 2458–2465.

Shafir S, Roughgarden J. 1996. The effect of memory length on individual fitness in a lizard. In: Belew RK, Mitchell M, editors. Adaptive individuals in evolving populations: models and algorithms. Reading (MA): Addison-Wesley. p. 173–181.

Shettleworth SJ. 1998. Cognition, evolution, and behavior. Oxford: Oxford University Press.

Shukitt-Hale B, McEwen JJ, Szprengiel A, Joseph JA. 2004. Effect of age on the radial arm water maze—a test of spatial learning and memory. Neurobiol Aging. 25:223–229.

Smulders TV, Dhondt AA. 1997. How much memory do tits need? Trends Ecol Evol. 12:417–418.

Spear NE. 1973. Retrieval of memory in animals. Psychol Rev. 80:163–194.

Spreng M, Rossier J, Schenk F. 2002. Spaced training facilitates long-term retention of place navigation in adult but not in adolescent rats. Behav Brain Res. 128:103–108.

Squire LR. 1989. On the course of forgetting in very long term memory. J Exp Psychol Learn Mem Cogn. 15:241–245.

Stephens DW. 1987. On economically tracking a variable environment. Theor Popul Biol. 32:15–25.

Stephens DW. 1991. Change, regularity, and value in the evolution of animal learning. Behav Ecol. 2:77–89.

Tomsic D, Romano A, Maldonado H. 1998. Behavioral and mechanistic bases of long-term habituation in the crab Chasmagnathus. In: Ehrlich YH, editor. Molecular and cellular mechanisms of neuronal plasticity. London: Kluwer Academic/Plenum Publishers. p. 17–35.

Williams GC. 1957. Pleiotropy, natural selection, and the evolution of senescence. Evolution. 11:398–411.

Wylie GR, Foxe JJ, Taylor TL. 2008. Forgetting as an active process: an fMRI investigation of item-method-directed forgetting. Cereb Cortex. 18:670–682.

Xia SZ, Liu L, Feng CH, Guo AK. 1997. Nutritional effects on operant visual learning in Drosophila melanogaster. Physiol Behav. 62:263–271.

Yanai S, Okaichi Y, Okaichi H. 2004. Long-term dietary restriction causes negative effects on cognitive functions in rats. Neurobiol Aging. 25:325–332.

Zelinski EM, Burnight KP. 1997. Sixteen-year longitudinal and time lag changes in memory and cognition in older adults. Psychol Aging. 12:503–513.