

1 **Information as a State Variable: A Tutorial** [Excerpted
2 from *Ectotherms in Changing Environments. Working in*
3 *Pasteur's Quadrant*, unpublished manuscript. Supported in part
4 by NSF grant 14- 51931 "OPUS: Ectotherms in Changing
5 Environments".]

6
7 February 24, 2021

8 Marc Mangel

9 Department of Biology, University of Bergen, Puget Sound Institute, University of Wash-
10 ington Tacoma, and Department of Applied Mathematics, University of California Santa
11 Cruz

12 marcmangel@protonmail.com

13
14 **Context**

15 Chapter 6 of the book focused on information as a state variable; this tutorial is an
16 self-contained, slightly expanded version of that chapter.

17 Understanding how organisms learn about the environment, other individuals, and
18 themselves will always be an important problem in biology. For example, in a recent
19 book, Barrett (2020) argued that the brain is a prediction machine focused on allostasis
20 (the problem of allocating across many different needs and functions) and thus needs to
21 learn constantly. In another recent book, Railsback and Harvey (2020) melded informa-
22 tion/learning and State Dependent Life History Theory (SDLHT, Mangel 2015, 2020) by
23 treating the dynamic optimization as a myopic process (*sensu* Mangel and Clark 1983).
24 Stamps and Bell (2020) make the point that unless one thinks about information properly

25 empirical studies in which the absence of a cue is considering a control may mislead both
26 the organism and the investigator.

27 As in the other tutorial (Mangel 2020a) and case study document (Mangel 2020b),
28 I refer to the canonical equations from Mangel (2015), although they are derived in the
29 tutorial concerning the path from demography to state dependent life history theory
30 (Mangel 2020).

31 Topics include

- 32 ● An introduction to the general questions that includes a roadmap for the tutorial.
- 33 ● Brief description of the rules of probability, including conditional probability, inde-
34 pendence, and Bayes's theorem – which underlies pretty much all of information
35 processing.
- 36 ● After that, we (you – the reader – and I) consider a variety of applications of Bayesian
37 ideas, in order of increasing complexity
 - 38 – Nesting male wrasse accepting or rejecting a potential satellite male.
 - 39 – Individuals learning about the quality of food in their environment.
 - 40 – The canonical equation for allocation with an uncertain end time.
 - 41 – Information and development towards one of two phenotypes in an uncertain
42 environment.
 - 43 – Long-term effects of early life experiences.
 - 44 .
- 45 ● We then turn to alternate ways of characterizing information, including
 - 46 – A sliding memory window, with application to superparasitism by the rose hips
47 fly.
 - 48 – Classical Bayesian approaches using the conjugate prior (if you do not know
49 what this means, keep reading because you will find out) including
 - 50 * The Poisson-Gamma model and the canonical equation for activity choice
51 when information is characterized by it.
 - 52 * The Beta-Binomial model and the canonical equation for allocation when
53 information is characterized by it.

54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71

* How threshold norms of reaction can emerge from Bayesian analysis.

- We next discuss modeling posteriors in a changing world.
- We then explore how the notion of entropy (which allows us to characterize distributions in a single number) allows us to characterize age-structured populations absent and with harvest and how an egg laying insect may maximize entropy as a response to its parasitoid.
- I provide a look to the future, summarizing some other topics in which Bayesian analysis is used and some interesting directions for research.
- In Appendix 1, I provide some more details for the calculation of entropy of the age-structured population. In Appendix 2, I discuss the Normal-normal Bayesian model and how it could be used by organisms for updating about temperature.

Sounds like a lot, but the pieces are digestible, so let's get going.

Introduction

All the business of war, and indeed all the business of life, is to find out what you don't know by what you do; that's what I called "guessing what was on the other side of the hill."
Sir Aurthur Wellsley, the Duke of Wellington

With the introduction of uncertainty - the fact of ignorance and necessity of acting upon opinion rather than knowledge - into this Eden-like situation [of no uncertainty], its character is completely changed. With uncertainty absent, man's energies are devoted altogether to doing things; it is doubtful whether intelligence itself would exist in such a situation; in a world so built that perfect knowledge was theoretically possible, it seems likely that all organic readjustments would become mechanical, all organisms automata. With uncertainty present, doing things, the actual execution of activity, becomes in a real sense a secondary part of life; the primary problem or function is deciding what to do and how to do it. (Knight 1921/1971, pg 268)

72 It is now well appreciated that organisms use information about their environment in
73 shaping life history decisions (Luttbeg and Schmitz 2000, McClinn and Stephens 2006,
74 McNamara and Dall 2010, Luttbeg and Trussell 2013, Trimmer and Houston 2014). The
75 theme of this tutorial is that physiological and informational states interact in potentially
76 complex ways but that the right framework will allow us to understand this interaction
77 (Mathot and Dall 2013). Including information in as a state variable in SDLHT models
78 has been conceptually easy for a long time (e.g. Krebs et al 1978, Mangel and Clark
79 1983), but sometimes leads to computational issues called the curse of dimensionality,
80 which arises when we have to track many possible outcomes. With modern computing
81 technology we are now able to consider problems of information that were inconceivable
82 even at the turn of the millennium (compare, for example Mangel and Clark (1983), Clark
83 and Mangel (2000) and Eliassen et al (2007, 2009)).

84 The behavioral ecology of information is a subject with plenty of potential pitfalls
85 (Schmidt et al 2010, Frankenhuis et al 2018). In changing environments, cues may be
86 inherently unreliable (Arnoldini et al. 2012) – environmental change may make previously
87 reliable cues less so (de Jong et al. 2010), or multiple cues may give conflicting information.
88 The key for resolving these problems is sampling and making use of the information so
89 obtained (Frankenhuis and Panchanathan 2011).

90 We will approach these problems using classical Bayesian analysis at the level of
91 Hilborn and Mangel (1997). There are other approaches. For example, McElreath (2016)
92 shows how to implement much of Bayesian analysis numerically, without using the math-
93 ematics of the classical approaches. I like this book very much! Higginson et al (2017)
94 show that treating physiological states as a source of information (i.e., how hungry the
95 organism is tells it something about food in the environment) is almost as effective as
96 Bayesian learning. I like this work very much too and encourage you to check both out.

97 **A Roadmap for This Tutorial**

98 Because at some points we will have to get into the mathematical forest, I want to give
99 you an overview of the path through the forest before we begin.

100 We will start with a reminder about the rules of probability at a simple level. This
101 will allow us to agree on things like independence, conditional probability, and Bayes
102 theorem. We will then flex your Bayesian methods with a number of applications: nesting
103 male ocellated wrasse *Symphodus ocellatus* accepting or rejecting a possible satellite male,
104 a laboratory experiment in which an individual receives fruit that might be sweet or

105 sour and updates the probability of sweetness, the canonical equation for allocation with
106 an uncertain end time, and two models of different complexity for the links between
107 information and development. These examples will show you how much can be done with
108 just the fundamentals of probability.

109 We will then discuss additional ways of characterizing information. The first is a
110 sliding memory window for experience, which we apply to the rose hips *Rhagoletis basilo*
111 fly laying eggs in a hip that has already been parasitized. The second is the classical
112 Bayesian approach of conjugate priors (if you do not know what this means now, do
113 not worry, you will learn in as the tutorial progresses) applied to learning the parameter
114 of a Poisson process or a binomial distribution. We will show how the former leads to
115 the negative binomial distribution and links to the canonical equation for activity choice
116 and how the latter links to the canonical equation for allocation. We will also show how
117 threshold norms of reaction emerge from Bayesian analysis and briefly discuss modern
118 Bayesian methods that are numerically intensive.

119 We then turn to ways of characterizing posterior distributions in changing worlds, by
120 either forgetting the updates parameters or by mixing the posterior and prior distributions
121 and I will show you a code for doing the latter.

122 Finally, we will turn to the population, rather than individual level, and discuss the
123 concept of entropy of a probability distribution and apply the idea to the simple model
124 of an age-structured population. We will close by linking back to behavior, discussing a
125 plant-seed herbivore-parasitoid system.

126 Shortly before I started my PhD at the University of British Columbia in 1974, I read
127 three wonderful articles by the statistician John Hammersley (Hammersley 1971, 1973,
128 1974) on the theme of how research is done in the mathematical sciences and the contri-
129 bution that mathematics can make to other sciences. These are beautiful and insightful
130 papers (especially for somebody coming into a theoretical PhD from an experimental
131 background, as I did). Over the years, without having re-read them carefully, I quoted
132 Hammersley as having said “The simpler the tool, the more likely it is to deliver the
133 goods” and “Scientists have come to expect every from mathematicians short of actual
134 help”. On re-reading the papers, I see that he never actually put it that way, but both
135 are wonderful epigrams!

136 The Rules of Probability

137 The the rules of probability are simple to derive but the devil is in the details of making
138 them operational. In this section, I will remind you of those rules.

- 139 • Our goal is to compute the probability of ‘events’ –things that can happen – which
140 at this point remain vaguely defined and are denoted by **A**, **B**, **C**; we let **S** denote
141 the collection of all possible things that can happen (Figure 1a).
- 142 • We let $\Pr\{\mathbf{A}\}$ denote the probability of the event **A** occurring and define it to be
143 the “area” of **A** divided by the “area” of **S**. I have put the word area in quotation
144 marks, because it is defining and interpreting the area in which those devilish details
145 appear. All other probabilities are defined similarly.
- Since something has to happen when we sum over all possible and mutually exclusive
(i.e. only one of them can happen) events

$$\Pr\{\mathbf{A}\} + \Pr\{\mathbf{B}\} + \Pr\{\mathbf{C}\} + \dots = 1 \quad (1)$$

- We use $\Pr\{\mathbf{A}, \mathbf{B}\}$ to denote the probability of events **A** and **B** both happening,
which will be determined from the ‘area’ common to both of them (refer to Fig 1a).
Then the probability that one of them occurs is

$$\Pr\{\mathbf{A} \text{ or } \mathbf{B}\} = \Pr\{\mathbf{A}\} + \Pr\{\mathbf{B}\} - \Pr\{\mathbf{A}, \mathbf{B}\} \quad (2)$$

146 We subtract $\Pr\{\mathbf{A}, \mathbf{B}\}$ on the right-hand side of Eqn 2 because otherwise it is
147 double-counted. If the events are mutually exclusive then $\Pr\{\mathbf{A}, \mathbf{B}\}=0$.

- We use $\Pr\{\mathbf{A} \mid \mathbf{B}\}$ to denote the probability of event **A** happening given that **B**
happens. The collection of all possible events is now **B**. By the definitions we have
adopted, as long as $\Pr\{\mathbf{B}\} > 0$

$$\Pr\{\mathbf{A} \mid \mathbf{B}\} = \frac{\Pr\{\mathbf{A}, \mathbf{B}\}}{\Pr\{\mathbf{B}\}} \quad (3)$$

- We say that two events are independent if knowing that one of them occurred does
not change the probability of the other occurring. That is **A** and **B** are independent

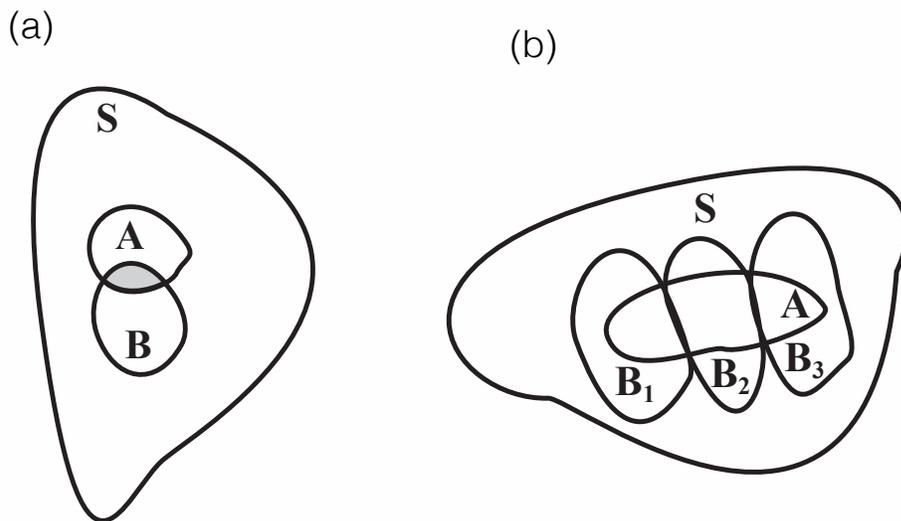


Figure 1: Simple figures allow us to understand abstract probability theory immediately and intuitively. a) In the collection of all things that can happen, S , two possibilities are A and B , which may have a commonality in that both of them happen (the shaded region). b) Here when A occurs, one of a number of mutually exclusive other events, here indicated by B_i may occur. This framework, in which the B_i are mutually exclusive sets up Bayes theorem.

if

$$\Pr\{A|B\} = \Pr\{A\} \tag{4}$$

Combining Eqns 3 and 4, we conclude that if two events are independent

$$\Pr\{A, B\} = \Pr\{A\}\Pr\{B\} \tag{5}$$

148
149
150

I cannot emphasize enough that Eqn 5 is not the definition of independent events, but it is the consequence of the definition, which is Eqn 4 (and totally logical, while Eqn 5 as a definition is totally mysterious).

We now rewrite Eqn 5 as

$$\Pr\{\mathbf{A}, \mathbf{B}\} = \Pr\{\mathbf{A}|\mathbf{B}\}\Pr\{\mathbf{B}\} \quad (6)$$

Since we can interchange the roles of \mathbf{A} and \mathbf{B} on the right hand side of this equation we can also write (with a bit of reorganization and remembering that if both \mathbf{A} and \mathbf{B} occur it does not matter in which order we write them)

$$\Pr\{\mathbf{B}|\mathbf{A}\} = \frac{\Pr\{\mathbf{A}, \mathbf{B}\}}{\Pr\{\mathbf{A}\}} \quad (7)$$

151 This equation only makes sense, of course, if the probability that \mathbf{A} happens is
152 greater than 0.

- This brings us to Bayes's theorem. Imagine a collection of mutually exclusive events $\mathbf{B}_1, \mathbf{B}_2, \dots, \mathbf{B}_I$ for which we know $\Pr\{\mathbf{A}|\mathbf{B}_i\}$ for each i (Figure 1b). If the \mathbf{B}_i represent all possible events that could occur when \mathbf{A} occurs, then it must be true that

$$\Pr\{\mathbf{A}\} = \sum_{i=1}^I \Pr\{\mathbf{A}, \mathbf{B}_i\} \quad (8)$$

153 We now flip this around and ask the question "Given that we know that \mathbf{A} occurred,
154 what is the probability that each of the different \mathbf{B}_i occurred?"

The answer to this question is found in the generalization of Eqns 7 and 8 and is

$$\Pr\{\mathbf{B}_i|\mathbf{A}\} = \frac{\Pr\{\mathbf{A}|\mathbf{B}_i\}\Pr\{\mathbf{B}_i\}}{\sum_{j=1}^I \Pr\{\mathbf{A}|\mathbf{B}_j\}\Pr\{\mathbf{B}_j\}} \quad (9)$$

155 Eqn 9 is called Bayes's theorem. Jaynes (2003) argues that Bayes theorem the logic of
156 science; it is particularly important in the 21st century (Taper and Ponciano 2016).

157 Let's unpack Eqn 9. First, note that in the denominator, I have summed over j rather
158 than i , to remind us that the denominator is independent of the numerator or the left
159 hand side of the equation. In fact, the denominator is $\Pr\{\mathbf{A}\}$, since each entry in the
160 summation is the probability of \mathbf{A} and \mathbf{B}_j occurring. Furthermore, the denominator is
161 a normalization constant, ensuring that the left-hand side is a probability distribution
162 summing to 1. Second, note how I have written the probability of \mathbf{A} and \mathbf{B}_i jointly
163 occurring as the conditional probability of \mathbf{A} given that \mathbf{B}_i occurred times the probability
164 that \mathbf{B}_i occurs. In Bayesian analysis, we call $\Pr\{\mathbf{B}_i\}$ the prior probability of \mathbf{B}_i and

165 $\Pr\{\mathbf{B}_i|\mathbf{A}\}$ the posterior probability of \mathbf{B}_i given that \mathbf{A} occurred.

166 You can find some more details about these derivations in Mangel (2006), pg 81-84,
167 examples of the counterintuitive nature of conditional probability in Hilborn and Mangel
168 (1997), pg 43-47, and examples of Bayesian analysis in Hilborn and Mangel (1997, pg
169 206ff).

170 Now that we have these rules, it is time to see how they are employed.

171 **Flexing Your Bayesian Muscles**

172 Armed with Eqns 7 and 9, we can do many interesting things (more complicated mathe-
173 matics will come later, if you are getting impatient).

174 **Nesting and Satellite Male Wrasse**

175 Among other things, Suzanne Alonzo works on nest guarding in ocellated wrasse *Sym-*
176 *phodus ocellatus* (Figure 2). In brief, the system can be described as follows (Alonzo and
177 Warner 1997, 2000; Stiver and Alonzo 2013, Alonzo and Heckman 2018). This wrasse
178 is found on rocky and seagrass habitat in shallow coastal waters and has separate sexes
179 with no sex change. The breeding season lasts for about 2 months and spawning is dem-
180 ersal, occurring daily. Individuals live about 2-3 years. Males (generally the larger ones)
181 construct nests and females visit many nests before spawning eggs that the male fertilize.
182 At the same time, smaller males – who could be helpers for the nesting male (e.g. in
183 guarding) – may also try to sneak matings. Such sneakers are able to fertilize eggs, but
184 do not participate in care or guarding of the eggs. Females prefer nests without sneaker
185 males.

We can simplify this situation to one in which satellite males are either classified as good (dependable, won't sneak matings) or bad (the reverse). Let us denote the prior probability that a potential satellite is good by p (think of it as the fraction of such males in the population before the nesting male has had any interactions). When they interact, the nesting male receives a cue from a potential satellite male that he is good, so let us introduce

$$p_g = \Pr\{\text{cue is that the potential helper is good} \mid \text{he actually is good}\}$$

$$p_b = \Pr\{\text{cue is that the potential helper is good} \mid \text{he actually is bad}\}$$

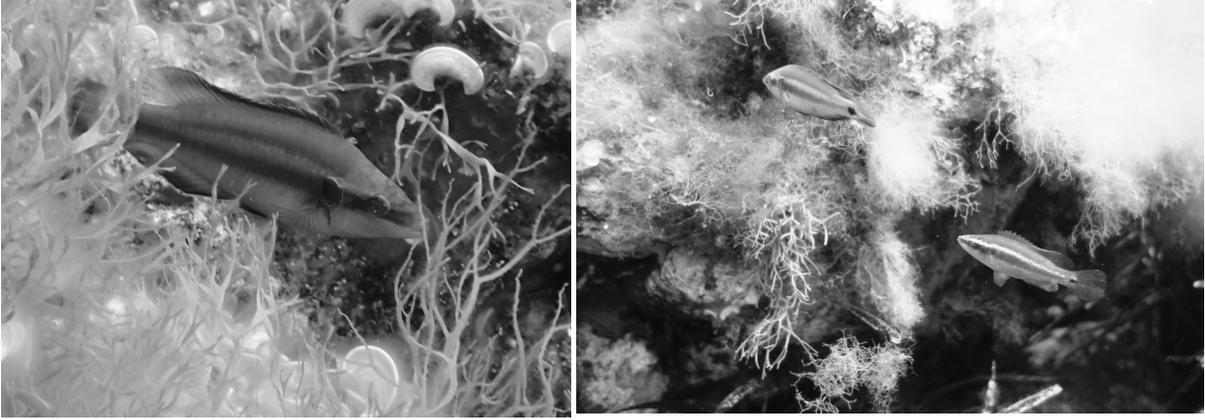


Figure 2: The ocellated wrasse *Symphodus ocellatus* studied by Suzanne Alonzo, in which males construct nests and guard eggs (left) and need to determine whether to accept or reject a possible helper at the nest (right).

186 That is, p_g is the probability that the nesting male has received an honest and informative
 187 cue and p_b is the probability that the nesting male has received a dishonest and non-
 188 informative cue.

Imagine now that a nesting male interacts with a potential satellite male and receives a cue that the potential satellite is good. We want to compute the probability $\Pr\{\text{satellite is actually good} \mid \text{good cue received}\}$. From Bayes's theorem this is

$$\Pr\{\text{satellite is actually good} \mid \text{good cue received}\} = \frac{\Pr\{\text{satellite is good and cue is good}\}}{\Pr\{\text{cue is good}\}}$$

The numerator in this expression is $p \cdot p_g$ since there is only one way for the potential helper to be truly good and give a good cue. But there are two ways to receive a good cue: 1) a good satellite gives a good cue, which happens with probability $p \cdot p_g$ or 2) a bad satellite gives a good cue, which happens with probability $(1 - p) \cdot p_b$. Since these are mutually exclusive events, we add them to determine that probability of receiving a good cue is $p \cdot p_g + (1 - p) \cdot p_b$. Thus after one interaction with this potential satellite in which a good cue is received, the posterior probability that this satellite is good is

$$p_1 = \frac{p \cdot p_g}{p \cdot p_g + (1 - p) \cdot p_b} \tag{10}$$

189 Let us imagine multiple interactions with the potential satellite. For definiteness, sup-
 190 pose that the cues received were good, bad, good, which we might write as the behavioral

191 sequence GBG (*sensu* Cane 1959). After the first cue of good is received, the probability
 192 that this potential satellite is a good one is no longer p but is p_1 given in Eqn 10. Hence,
 193 we use p_1 – the posterior probability that the potential male is good – as the prior in
 194 Bayes’s theorem to update when the nesting male receives the second cue.

Since the second cue of our hypothetical sequence is bad, we are asking for the proba-
 bility that the male is good given that a cue that he is bad is received. Hence the analogy
 of Eqn 10 is

$$p_2 = \frac{p_1 \cdot (1 - p_g)}{p_1 \cdot (1 - p_g) + (1 - p_1) \cdot p_b} \quad (11)$$

195 The numerator in Eqn 11 is the probability that the potential satellite is good and a bad
 196 signal is recieved from him and the denominator is the probability that a bad signal is
 197 received. We use p_2 as the prior when updating after the the third cue.

It is clear that we can write these equations in a more general form. Suppose that p_n
 is the current estimate that this potential satellite is good. Then if the next cue is that
 he is good, the posterior is

$$p_{n+1} = \frac{p_n \cdot p_g}{p_n \cdot p_g + (1 - p_n) \cdot p_b} \quad (12)$$

and if the next cue is that he is bad, the posterior is

$$p_{n+1} = \frac{p_n \cdot (1 - p_g)}{p_n \cdot (1 - p_g) + (1 - p_n) \cdot p_b} \quad (13)$$

198 To illustrate these equations, let us consider the informational status of the nesting male
 199 after five encounters with the same potential satellite male. The cues received during these
 200 encounters can be GGGGG, GBGGG, GGBGG, GGGBG, GGGGB, GBBGG, GBGBG,
 201 GGGBB, etc. With 5 cues, there are $2^5 = 32$ possible sequences of cues. The posterior p_5
 202 will be bounded above by that associated with the cue GGGGG and below by BBBB,
 203 which are computed from Eqn 12 or 13 respectively.

204 In Figure 3, I show the posterior probabilities for 5 cases: the boundaries of GGGGG
 205 and BBBB and intermediate cases of GBGGG, GGBGG (one bad signal) and GBBGG
 206 (two bad signals), for these parameters: $p(0) = 0.3, p_g = 0.8, p_b = 0.5$. We can learn
 207 much from this figure. For example, starting at the prior probability, cues lead to a
 208 fanning out of the posterior, with the boundaries of the fan set by GGGGG and BBBB.

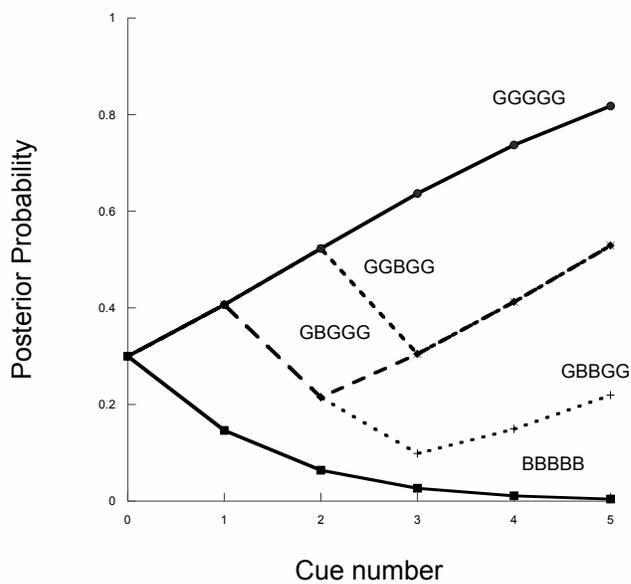


Figure 3: The posterior probability that a potential helper will be good, given the sequence of cues, in which all helpers have the same prior probability of being good. The cue sequences BBBBB and GGGGG bound the region of posterior probability. Judy Stamps called this “fanning out”.

209 The interior of this fan will not be completely filled, since for 5 cues there are only 32
 210 possible trajectories, but a variety of trajectories will fill the fan. Trajectories with the
 211 same number of cues of B will end in the same spot, as illustrated by the GBGGG and
 212 GGBGG trajectories; this is a result of the Markov property of the Bayesian model – the
 213 future value only depends upon the current value and not the history that got us to the
 214 current value.

215 Female wrasse also face informational problems (e.g. Luttbeg and Warner 1999) but
 216 we will not get into them here.

217 Flexing Your Bayesian Muscles: Sweet Fruit in the Laboratory

218 Here’s another situation, suggested to me by Judy Stamps (for an example of chimpanzees
 219 in the field, see Dominy et al (2016)). Instead of a range of possible cues and a fixed prior,
 220 we might have a single set of cues but a range of possible prior values. Such a situation
 221 could occur for example when we bring animals into the laboratory from a natural setting
 222 and give them exactly the same set of cues. For example, imagine that in nature fruit
 223 comes in two colors, Red (R) and Yellow (Y) and there is variation in sweetness in the
 224 red fruit either annually or by location, but when it is sweet, a red fruit it is much sweeter

225 than a yellow fruit and thereby preferred. The behavioral problem, which you may want
226 to model, for the foraging individual is then whether to spend time handling a yellow fruit
227 when encountered or skip it and keep looking for red fruit.

228 Thus, an individual has an initial probability $p(0)$ that red fruit is sweet, but this varies
229 across individuals because of previous experience. We now offer individuals a sequence of
230 fruit, say 15 sweet fruit in a row, and ask how the posterior distributions evolve in time.
231 We can use Eqn 12 with reinterpretation of p_g as the probability that a red fruit is sweet
232 and p_b that the red fruit is sour.

233 In this case, the posterior distributions ‘fans in’ towards the conclusion that red fruit
234 this year or in this location is sweet (Figure 4); of course we know the answer but the
235 animal does not. The pace at which this conclusion is reached depends upon the prior,
236 but after a sufficient number of red fruit that are sweet, the prior information is swamped
237 by experience. This figure illustrates the point that once we have enough experience, the
238 prior becomes unimportant. Which raises the question, how much is enough - that is
239 what is the value of additional information given our current uncertainty? We will come
240 back to this question. As an exercise, you might redo the calculations when we give the
241 animals sweet red fruit and sour red fruit alternately, so that after 15 fruit, the experience
242 will be either 7 sweet and 8 sour or the reverse. How do you expect the fanning in to
243 look?

244 Where do the priors come from? In our examples, by the time a male has reached
245 nesting status, he may have encountered and observed a large number of satellite males
246 (or acted as one himself) so that he can construct a prior from his own experience; animals
247 encountering fruit in the wild will have experienced a mixture of sweet and sour red fruit.
248 In other situations, we can imagine that the individuals have a prior given to them by
249 their mother – either through the genotype or other maternal signals.

250
251 There is controversy concerning Bayesian priors, which are often been called “sub-
252 jective” and have lead to great fights between statisticians. I believe that scientists are
253 facultative Bayesians who will bring the right tool to the job (also see Efron (1986) and
254 the discussion following it by Hermann Chernoff). If all you need to know is if two means
255 are different, then a simple t-test will usually do the job and you don’t need Bayesian
256 statistics, but other times you do.

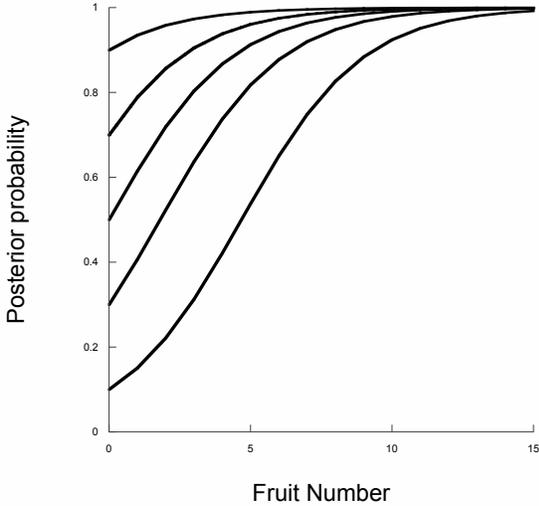


Figure 4: The alternative to “fanning out” is “fanning in”, in which individuals have different prior probabilities (in this case for whether a red fruit is sweet or not) but receive the same sequences of cues, as they could in a controlled laboratory experiment.

257 **Flexing Your Bayesian Muscles: The Canonical Equation for Allocation with**
 258 **an Uncertain End Time**

Let’s return to the simplest version of the canonical equation for allocation in a pro-ovigenic insect

$$F(x, s) = e^{-m}(1 - \eta)F(x, s + 1) + \eta \max_{c \leq x} [f(c) + e^{-m(c)} F(x - c, s + \tau(c))] \quad (14)$$

where x is the number of eggs remaining at time s , m is the rate of mortality while searching, η is the probability of encountering a host in a single period of search, c is the clutch laid, $f(c)$ is the increment in lifetime fitness when she lays a clutch of size c and $\tau(c)$ is the time taken to lay that clutch.

Previously we assumed a fixed end time S with $F(x, S) = 0$ for all x ; for an insect that lives a single year, we interpreted S as the time of the first frost, assuming that the frost was fatal. Suppose, however, that the time of the first frost varies from year to year and we treat it as a random variable \tilde{S} with a probability distribution at the start of the season given by

$$\phi(s) = \text{Probability that the first frost is at time } s = \text{Prob}\{\tilde{S} = s\} \quad (15)$$

259 We expect that $\phi(s)$ will be 0 for times less than the earliest time s_E for the frost and for
 260 times greater than the latest time of the frost.

At time s , the parasitoid can continue to look for hosts as long as the season ends after period s , so we need to compute the probability that the season ends exactly at period s , given that it ends at s or later

$$\Phi(s) = \text{Prob}\{\tilde{S} = s | \tilde{S} \geq s\} \quad (16)$$

which we compute from Bayes theorem (also see Mangel and Clark (1988), pg 72-73)

$$\Phi(s) = \frac{\phi(s)}{\sum_{s'=s}^{s'=s_L} \phi(s')} \quad (17)$$

261 Information is captured in $\Phi(s)$, which in principle can change in every time period.

Once we have $\Phi(s)$, we modify Eqn 14 to account for the uncertain end time and that the parasitoid will be alive after period s with probability $1 - \Phi(s)$

$$F(x, s) = e^{-m}(1 - \eta)(1 - \Phi(s))F(x, s + 1) + \eta \max_{c \leq x} [f(c) + e^{-m(c)}(1 - \Phi(s))F(x - c, s + \tau(c))] \quad (18)$$

262 I suggest that you try this out, by choosing $\phi(s)$ and modifying the code that you've
 263 written for the canonical equation for resource allocation.

264 Information and Development

265 One of the classical interpretations of development (Waddington 1957) is that of a landscape
 266 (Figure 5) in which an organism starts in an undifferentiated state (the top of the hill)
 267 and may end up in one of a number of fully differentiated states (the bottom of the hill,
 268 here with four possibilities shown). A figure such as this begs the question ‘what sets the
 269 ball rolling down the hill on a particular trajectory?’. The top of the hill and each of the
 270 points where the path can bifurcate is an unstable state and random fluctuations may
 271 drive the ball in one direction or another¹.

272 How might an organism steer itself along the possible paths, given the information
 273 that it has received thus far in life in the presence of random fluctuations? In this section,

¹These ideas also apply in population ecology. Such an interpretation, for example, has been given to the classic experiments of Thomas Park on flour beetles *Tribolium* spp; see Mangel and Ludwig (1977) and Mangel (1994d) for both citations to the flour beetle work and some of the stochastic methods.

274 we explore work that has addressed this question. We use very simple Bayesian updating,
 275 similar to the examples we have done thus far, and will see that Bayesian methods can
 276 play an important role in the understanding of development. Additional entries to the
 277 literature beyond what we do are Nettle et al (2013), Stamps and Krishnan (2014), Nettle
 278 and Bateson (2015), and Stamps and Frankenhuis (2016).

279 **Developing Towards One of Two Phenotypes**

280 As always, it helps to develop intuition with ideas that are not fixed to a specific system
 281 but have much in common with many systems. Frankenhuis and Panchanathan (2011)
 282 and Panchanathan and Frankenhuis (2016) have a particularly lovely model.

We assume that after birth organisms disperse and settle into one of two different environments, E_0 or so E_1 (e.g. predator poor or predator rich), in which they develop, reproduce, and die. We assume the organisms are semelparous and reproduction depends on how well their phenotype is matched to the environment in which they settled . We let $p_0 = \text{Prob}\{\text{settling in environment } E_0 \text{ after dispersal}\}$; the probability of being in environment E_1 after dispersal is $1 - p_0$. We assume, but do not specify the details, that there is an optimal phenotype P_i for each environment (e.g. armored or highly mobile in the predator rich state, with a cost to maintaining defense, or escape mechanisms that are not needed in the predator poor environment).

Information enters by allowing the focal individual to only imperfectly know the environment through cues

$$p(c_0|E_0) = \Pr\{\text{cue that the environment is } E_0 \mid \text{Environment is } E_0\} \quad (19)$$

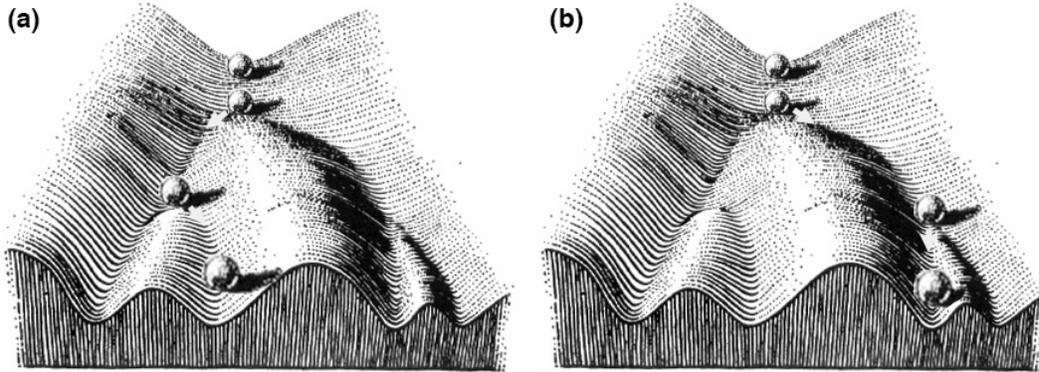
$$p(c_1|E_1) = \Pr\{\text{cue that the environment is } E_1 \mid \text{Environment is } E_1\} \quad (20)$$

For simplicity, assume that the cues are symmetric in the sense that $p(c_0|E_0) = p(c_1|E_1)$. After some amount of development time, the organism will have accumulated data $D(x_0, x_1)$ that consists of x_0 cues of environment E_0 and x_1 cues of environment E_1 . The probability of these data, conditioned on the environment is given by the binomial distribution

$$P(D(x_0, x_1)|E_0) = \binom{x_0 + x_1}{x_0} p(c_0|E_0)^{x_0} (1 - p(c_0|E_0))^{x_1} \quad (21)$$

$$P(D(x_0, x_1)|E_1) = \binom{x_0 + x_1}{x_1} p(c_1|E_1)^{x_1} (1 - p(c_1|E_1))^{x_0} \quad (22)$$

Undifferentiated



Developed

Figure 5: Waddington's (1957) vision of the developmental landscape as a process combining deterministic (the ball rolling downhill) and stochastic (which valley it enters when encountering bifurcating paths).

We use Bayes's theorem to obtain posteriors for $p_0 = \Pr\{\text{environment is } E_0 | D(x_0, x_1)\}$ and $p_1 = \Pr\{\text{environment is } E_1 | D(x_0, x_1)\}$,

$$p'_0(D(x_0, x_1)) = \frac{P(D(x_0, x_1) | E_0) p_0}{P(D(x_0, x_1) | E_0) p_0 + P(D(x_0, x_1) | E_1) p_1} \quad (23)$$

$$p'_1(D(x_0, x_1)) = \frac{P(D(x_0, x_1) | E_1) p_1}{P(D(x_0, x_1) | E_0) p_0 + P(D(x_0, x_1) | E_1) p_1} \quad (24)$$

To model development, we assume that development must be completed by time S , when fitness is assessed. At that time, the phenotype consists of $Y_0(S)$ developmental steps suited for environment E_0 and $Y_1(S)$ developmental steps suited for environment E_1 . Frankenhuis and Panchanathan (2011) consider three kinds of terminal fitness functions, with linear, diminishing, and increasing returns according to phenotype. These are

respectively, with $y_0 = Y_0(S)$ and $y_1 = Y_1(S)$,

$$\phi(x_0, x_1, y_0, y_1) = p'_0(D(x_0, x_1))y_0 + p'_1(D(x_0, x_1))y_1 \quad (25)$$

$$\phi(x_0, x_1, y_0, y_1) = p'_0(D(x_0, x_1))\alpha(1 - e^{-\beta y_0}) + p'_1(D(x_0, x_1))\alpha(1 - e^{-\beta y_1}) \quad (26)$$

$$\phi(x_0, x_1, y_0, y_1) = p'_0(D(x_0, x_1))\alpha(e^{\beta y_0} - 1) + p'_1(D(x_0, x_1))\alpha(e^{\beta y_1} - 1) \quad (27)$$

where α and β , both greater than 0, are parameters that measure the strength and shape of the dependence of the terminal fitness function on the developmental progress.

Previous to the terminal time, the organism can choose between sampling for a cue of the state of the world, adding an increment towards a phenotype suitable for E_0 , or adding an increment towards a phenotype suitable for E_1 . We define

$$F(x_0, x_1, y_0, y_1, s) = \text{Maximum expected value of } \phi(X_0(S), X_1(S), Y_0(S), Y_1(S)) \\ \text{given that } X_0(s) = x_0, X_1(s) = x_1, Y_0(s) = y_0, Y_1(s) = y_1 \} \quad (28)$$

283 Thus, our fitness function combines information, summarized by (x_0, x_1) , and develop-
284 ment, summarized by (y_0, y_1) .

We let $V_{sample}(x_0, x_1, y_0, y_1, s)$, $V_0(x_0, x_1, y_0, y_1, s)$, and $V_1(x_0, x_1, y_0, y_1, s)$ denote the fitness value of sampling at time s , developing towards the phenotype appropriate for E_0 , and developing towards the phenotype appropriate for E_1 respectively. Then

$$F(x_0, x_1, y_0, y_1, s) = \max[V_{sample}(x_0, x_1, y_0, y_1, s), \\ V_0(x_0, x_1, y_0, y_1, s), V_1(x_0, x_1, y_0, y_1, s)] \quad (29)$$

In order to compute the fitness value of sampling, we need to know the probabilities $p_s(c_0), p_s(c_1)$ of obtaining cues of environments E_0 and E_1 respectively. To obtain them, we combine the probability of getting a cue, conditioned on the environment, and the current posterior probability of the relevant environment to obtain

$$p_s(c_0) = p(c_0|E_0)p'_0(D(x_0, x_1)) + p(c_0|E_1)p'_1(D(x_0, x_1)) \\ p_s(c_1) = p(c_1|E_0)p'_0(D(x_0, x_1)) + p(c_1|E_1)p'_1(D(x_0, x_1)) \quad (30)$$

so that the fitness values of the three choices are

$$\begin{aligned} V_{sample}(x_0, x_1, y_0, y_1, s) &= p_s(c_0)F(x_0 + 1, x_1, y_0, y_1, s + 1) \\ &\quad + p_s(c_1)F(x_0, x_1 + 1, y_0, y_1, s + 1) \\ V_0(x_0, x_1, y_0, y_1, s) &= F(x_0, x_1, y_0 + 1, y_1, s + 1) \\ V_1(x_0, x_1, y_0, y_1, s) &= F(x_0, x_1, y_0, y_1 + 1, s + 1) \end{aligned} \tag{31}$$

285 Thus, V_{sample} involves a change in informational state but not developmental state; V_0
286 and V_1 involved a change in developmental state but not informational state.

287 Although this model does not include mortality or physiological state (both of which
288 we will include with the next model), implementing it is not a trivial matter. We have to
289 specify the prior probability of ending in environment E_0 after dispersal, the probability
290 that the cue is accurate, the form of the fitness function, and the parameters that go into
291 the fitness functions in Eqns 26 and 27. See the paper of Frankenhuis and Panchanathan
292 (2011) for those details.

293 Two insights from the resulting computations are i) the emergence of critical windows
294 for development and ii) individual differences in plasticity. Frankenhuis and Panchanathan
295 (2011) found that when cues are either highly informative or close to non-informative, nat-
296 ural selection is predicted to lead to behaviors in which individuals sample minimally or
297 not at all. In such a case, development will not be a function of the information that an
298 individual receives. On the other hand, when cues are moderately informative individ-
299 uals may sample extensively. But because cues are stochastic, some individuals achieve
300 sufficient certainty about the environment early in the sampling process to complete de-
301 velopment and others only late in the sampling process. This will lead to individual
302 differences of the timing of development – individuals are predicted to stop sampling the
303 environment at different times.

304 **Long-term Effects of Early Life Experiences**

305 We now turn to a model (English et al 2016) that includes growth and mortality and which
306 shows how the use of information during growth before sexual maturity can generate
307 apparent windows for development, as happens with the salmonids (e.g. Mangel 1994a,
308 Thorpe et al 1998, Satterthwaite et al 2009).

As above, we assume that after birth individuals disperse randomly to a patchy world
where environmental quality varies; once in a environment, individuals remain to grow,

reproduce, and die there. For simplicity, we again assume only two kinds of environments. One is higher quality in every way: reproduction as a function of size is greater, the probability of finding food as a function of foraging effort is higher, and the risk of mortality as a function of foraging effort is lower. In particular, if α is the fraction of a day spent foraging, then the probability of finding food and probability of mortality when in environment E_i are

$$\begin{aligned}\lambda(\alpha|E_i) &= \alpha\beta_i \\ \mu(\alpha|E_i) &= \mu_0(1 + \mu_i\alpha^\gamma)\end{aligned}\tag{32}$$

309 where $\beta_1 > \beta_2$, μ_0 is a baseline level of mortality common to both kinds of patches, $\gamma > 1$ is
 310 a shape parameter relating foraging effort and mortality and $\mu_1 < \mu_2$. Clearly, individuals
 311 will prefer to be in environment type 1. In the Appendix to their paper, English et al
 312 (2016) consider the cases in which $\mu_1 = \mu_2$ and $\mu_1 > \mu_2$, the latter corresponding to the
 313 classic patch selection problem.

314 To focus on the informational aspects of the problem, English et al (2016) use a simple
 315 growth model. If $X(s)$ is size at time s and food is found then $X(s + 1) = X(s) + 1$;
 316 otherwise $X(s + 1) = X(s)$.

The realized reproduction for an individual of size x in environment i is assumed to be a sigmoidal function of size

$$\Phi(x|E_i) = \frac{1}{a_i + e^{-b_i(x-c_i)}}\tag{33}$$

317 where a_i, b_i and c_i , all greater than zero, are parameters. When x is sufficiently large
 318 $\Phi(x|E_i) \approx \frac{1}{a_i}$ so a_i characterizes maximum reproduction. When $x = c_i$, $\Phi(x|E_i) = \frac{1}{a_i+1}$,
 319 giving a connection between a_i and c_i . Finally b_i , which English et al (2016) treat as
 320 independent across environments, characterizes the dispersion of the sigmoidal shape –
 321 basically how knife-edged the reproductive function is, with sharper transitions around
 322 $x = c_i$ as b_i increases (if that is not clear to you, spend some time making graphs of
 323 $f(x) = \frac{1}{1+e^{-b(x-3)}}$ and let b vary between 0.5 and 5).

We let p denote the current probability that the organism is in environment is E_1 and

define fitness by

$$F(x, p) = \text{Maximum expected reproductive success} \mid \{ X(s) = x \text{ and} \\ \text{the estimate that the current environment is } E_1 = p \} \quad (34)$$

324

Note that there is no terminal time in this definition. We will tackle this issue shortly.

The organism has two choices given its current size and estimate of the environment: i) to continue to grow and update the current estimate of the environment depending on whether it succeeded with foraging or not, ii) or to mature at its current size, claim the associated reproductive success and die but avoid the risk of dying before reproducing. Thus we write

$$F(x, p) = \max \{ V_{grow}(x, p), V_{mature}(x, p) \} \quad (35)$$

where $V_{grow}(x, p)$ and $V_{mature}(x, p)$ are respectively the fitness values of growing and maturing given the current physiological and informational states.

The value of maturing is easier to compute, so let's tackle that first. Since the reproductive success of an individual if it is in environment E_i is $\Phi(x|E_i)$ given by Eqn 33 and the probability that it is in environment 1 is p we have

$$V_{mature}(x, p) = p\Phi(x|E_1) + (1 - p)\Phi(x|E_2) \quad (36)$$

When we compute the value of growing, there is also an optimal foraging decision to be made: the tradeoff between foraging success and mortality implied by Eqn 32. If we denote the updated probability of being in environment E_1 when food is found by p_s (for successful foraging) and by p_u (for unsuccessful foraging) when food is not found then

$$V_{grow}(x, p) = \max_{\alpha} \left(p(1 - \mu(\alpha|E_1))[\lambda(\alpha|E_1)F(x + 1, p_s) \right. \\ \left. + (1 - \lambda(\alpha|E_1))F(x, p_u)] \right. \\ \left. + (1 - p)(1 - \mu(\alpha|E_2))[\lambda(\alpha|E_2)F(x + 1, p_s) \right. \\ \left. + (1 - \lambda(\alpha|E_2))F(x, p_u)] \right) \quad (37)$$

We determine the updated probabilities via Bayes theorem

$$\begin{aligned}
 p_s &= \frac{p\lambda(\alpha|E_1)(1 - \mu(\alpha|E_1))}{p\lambda(\alpha|E_1)(1 - \mu(\alpha|E_1)) + (1 - p)\lambda(\alpha|E_2)(1 - \mu(\alpha|E_2))} \\
 p_u &= \frac{p(1 - \lambda(\alpha|E_1))(1 - \mu(\alpha|E_1))}{p(1 - \lambda(\alpha|E_1))(1 - \mu(\alpha|E_1)) + (1 - p)(1 - \lambda(\alpha|E_2))(1 - \mu(\alpha|E_2))}
 \end{aligned} \tag{38}$$

We still have the sticky problem of how to solve the SDP since there is no time variable in it, and hence no formal end condition. The solution is to use a method called value iteration (e.g. Puterman 1994). To do this, we introduce a series of fitness functions $F^n(x, p)$, $n = 0, 1, 2, \dots$, with the upper limit of n explained momentarily. We then write the SDP equation as

$$F^{n+1}(x, p) = \max \{ V_{grow}^n(x, p), V_{mature}^n(x, p) \} \tag{39}$$

325 where the fitness values of continuing to grow or to mature are computed from Eqn 36
 326 and 37 using $F^n(x, p)$. Once we specify $F^0(x, p)$, we are off to the races. There are general
 327 theorems (e.g. Puterman 1994) that as n increases, $F^n(x, p) \rightarrow F(x, p)$, the solution of
 328 Eqn 35. However, we cannot let n increase indefinitely, so I need a stopping rule such as
 329 $\max |F^n(x, p) - F(x, p)| < \epsilon$, where ϵ is a predetermined cut-off. English et al (2016) used
 330 $\epsilon = 10^{-6}$, and found that it was reached for n in the range of 260-325, depending upon
 331 assumptions on how predation and foraging covary. To get this method working, we still
 332 need to specify $F^0(x, p)$. English et al (2016) randomly assigned values. In my work, I
 333 have sometimes simply set it to 0. However, if we recognize that value iteration implicitly
 334 introduces a time like variable, then $n = 0$ is essentially like the end condition (since in
 335 value iteration n increases while in the models we have done thus far time decreases).
 336 Hence, a good choice is $F^0(x, p) = p\Phi(x|E_1) + (1 - p)\Phi(x|E_2)$.

Now, we can define the value of information² associated with any state by computing the difference between fitness when updating occurs, as we have done above, and fitness without updating, which would be Eqn 37 with p_s and p_u replaced by p – regardless of what happens, the animal still uses the initial value of the probability. If we let $\tilde{F}(x, p)$ and α^* denote respectively fitness without updating information and then the value of

²Pike et al (2016) provide general analytical expression for the reproductive (i.e. fitness) value of information. Their paper is a bit technical, but worth the effort.

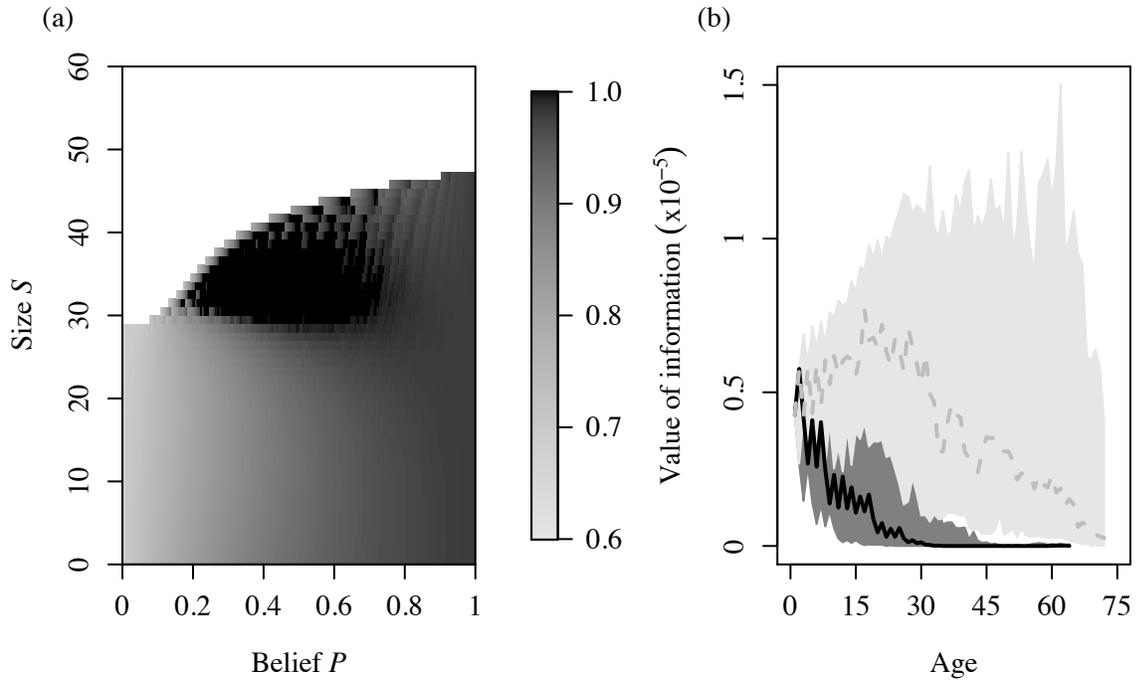


Figure 6: Optimal foraging effort, $\alpha^*(x, p)$ and boundary between maturing (gray) and continuing to grow, as a function of the current probability p of being in environment E_1 and current state x . b) Value of information (median and interquartile range) for 100 individuals developing in the environment E_1 (solid line, blue shading) or E_2 (dashed line, orange shading) environment. From English et al (2016).

information is

$$\begin{aligned} V_{info}(x, p) = & [p\lambda(\alpha^*|E_1)(1 - \mu(\alpha^*|E_1) + \\ & (1 - p)\lambda(\alpha^*|E_2)(1 - \mu(\alpha^*|E_2))][F(x, p_s) - \tilde{F}(x, p)] \\ & + [p(1 - \lambda(\alpha^*|E_1))(1 - \mu(\alpha^*|E_1) + \\ & (1 - p)(1 - \lambda(\alpha^*|E_2))(1 - \mu(\alpha^*|E_2))][F(x, p_u) - \tilde{F}(x, p)] \end{aligned} \quad (40)$$

337 The terms on the right hand side of this equation are respectively the probability of
338 surviving and finding food times the difference in fitness when updating accounts for
339 finding food and the probability of surviving and not finding food times the difference
340 in fitness when updating accounts for not finding food. In Figure 6, I show some of the
341 results of English et al (2016) on foraging effort as a function of age and size and the
342 value of information as a function of age. Their paper has many other interesting results
343 and is well worth your time and effort.

344 Additional Ways of Characterizing Information

345 We now move beyond the simplest case of the environment being just one of two states.

346 A Sliding Memory Window

347 We begin with a simple model for information that Bernie Roitberg and I used when try-
348 ing to understand the circumstances under which an apple maggot *Rhagoletis pomonella*
349 Walsh was predicted to oviposit in a fruit that had already received an egg and had been
350 marked with a pheromone by a different female (Mangel and Roitberg 1989). *R. Basiola*
351 lays eggs in clumps of roses that are discretely separated in space. Our goal was to predict
352 when a focal female would accept a previously parasitized and marked host, as a function
353 of time since the last oviposition by the focal female. When she lays an egg in a previously
354 parasitized host, she is said to have super-parasitized.

355 Imagine that the fly keeps track of what happened in the last N searches and let O_n
356 = 1 if the outcome was that a host was found and $O_n = 0$ if no host was found, where O_1
357 is the most recent search attempt and O_N is the most distant (in time) search attempt.

A simple estimate for the probability of encountering a host on the next search is

$$\hat{p} = \frac{1}{N} \sum_{n=1}^N O_n \quad (41)$$

358 We called this a sliding memory model because after each search, the most distant outcome
 359 is completely forgotten as a new outcome is added. Eqn 41 is a method of moments
 360 estimator (i.e. the average), now out of favor but once – before computing was so easy –
 361 quite popular.

362 We thus have a vector of information $\mathbf{I}(s) = \{O_1, O_2, O_3, \dots, O_N\}$ in which information
 363 about past host encounters slides along until it is completely forgotten. At any time,
 364 then, the current vector of information is $\mathbf{I}(s) = \mathbf{i}$ from which we can create an estimate
 365 of finding a host in the next search using Eqn 41.

This sliding memory model can be modified so that recent events are more valuable in
 computing the probability of encountering a host than past ones by introducing a memory
 weighting factor $w \leq 1$ and replacing Eqn 41 by

$$\hat{p} = \frac{O_1 + wO_2 + w^2O_3 + \dots + w^{N-1}O_N}{1 + w + w^2 + \dots + w^{N-1}} \quad (42)$$

The denominator in this equation is the geometric series and can be simplified because
 (Appendix 1)

$$1 + w + w^2 + \dots + w^{N-1} = (1 - w^N)/(1 - w) \quad (43)$$

so that Eqn 42 becomes

$$\hat{p} = \frac{(O_1 + wO_2 + w^2O_3 + \dots + w^{N-1}O_N)(1 - w)}{1 - w^N} \quad (44)$$

This result allows us to treat the estimate in Eqn 42 as a state variable $P(\mathbf{i}, s)$. A little
 bit of algebra shows that if O_{next} is the outcome of the next search then (Mangel and
 Roitberg 1989 for details)

$$P(\mathbf{i}', s + 1) = O_{next}(1 - w)/(1 - w^{N+1}) + wP(\mathbf{i}, s)(1 - w^N)/(1 - w^{N+1}) \quad (45)$$

366 where \mathbf{i}' is the updated informational state based on the next encounter. This simple
 367 sliding memory model makes information a dynamic state variable.

368 In the experiments (Mangel and Roitberg, 1989), we released a fly in a field cage

369 with a tree of about 1 m³ canopy volume holding between 8 and 64 individually iden-
370 tified fruit. After the first two ovipositions, we set the clock measuring the time since
371 last oviposition to 0, and followed the her encounters (unparasitized or previously para-
372 sitized host) and oviposition behavior. Thus the data consist of the Time Since the Last
373 Oviposition (TSLO), the fraction of previously parasitized hosts (FRAC) among the last
374 5 hosts encountered (doing a sliding memory model more than 30 years ago was limited
375 by computational power!) and whether a previously parasitized host was accepted or
376 rejected (an unparasitized host is predicted to receive an oviposition, regardless of TSLO
377 and FRAC, and almost always did so).

378 For the details of the SDP model see Mangel and Roitberg (1989). In brief, its structure
379 is: i) when no host is encountered, the female needs to survive to the next period and egg
380 complement does not change; ii) when an unparasitized hosts is encountered, the female
381 oviposits, giving an increase in lifetime fitness by f_u and decreasing egg complement by 1;
382 iii) when a previously parasitized host is encountered, the female compares future fitness
383 if she rejects the host or if she oviposits thus reducing egg complement and increasing
384 accumulated lifetime fitness by $f_p < f_u$.

385 Intuition suggests that as TSLO increases, the fly will be more likely to accept a
386 previously parasitized host at lower values of FRAC (Figure 7a – where the line is drawn
387 using the theory of Iwasa et al (1984)). However, only about half of the observations of
388 oviposition in a previously parasitized fruit fall in the “Accept” region of the TSLO/FRAC
389 plane (Figure 7b).

390 When the weighting parameter w in Eqn 42 is less than 1, the same values of FRAC
391 will correspond to different estimates of the probability of encountering an unparasitized
392 host. That is if FRAC = 0.4, 2 of the last 5 hosts encountered were previously parasitized
393 but the sequences PPUUU and UUUPP provide different views about the current state
394 of the world. Including this recognition leads us to divide the TSLO/FRAC plane into
395 three regions: one in which a previously parasitized host is accepted, one in which it is
396 rejected, and one in which it is sometimes rejected (Figure 7c).

397 When we compare the the theory that takes into account the weighting of recency of
398 encounters (Figure 7d), we discover that instances in which a marked fruit was accepted
399 for oviposition (indicated by a star). I also show fifteen quasi-randomly selected instances
400 from Mangel and Roitberg (1989) in which a marked fruit was rejected (circles). Note
401 that nearly all acceptances fall in the “Accept” or “Maybe” regions and that the rejections
402 fall in the ”Maybe” and ”Reject” regions.

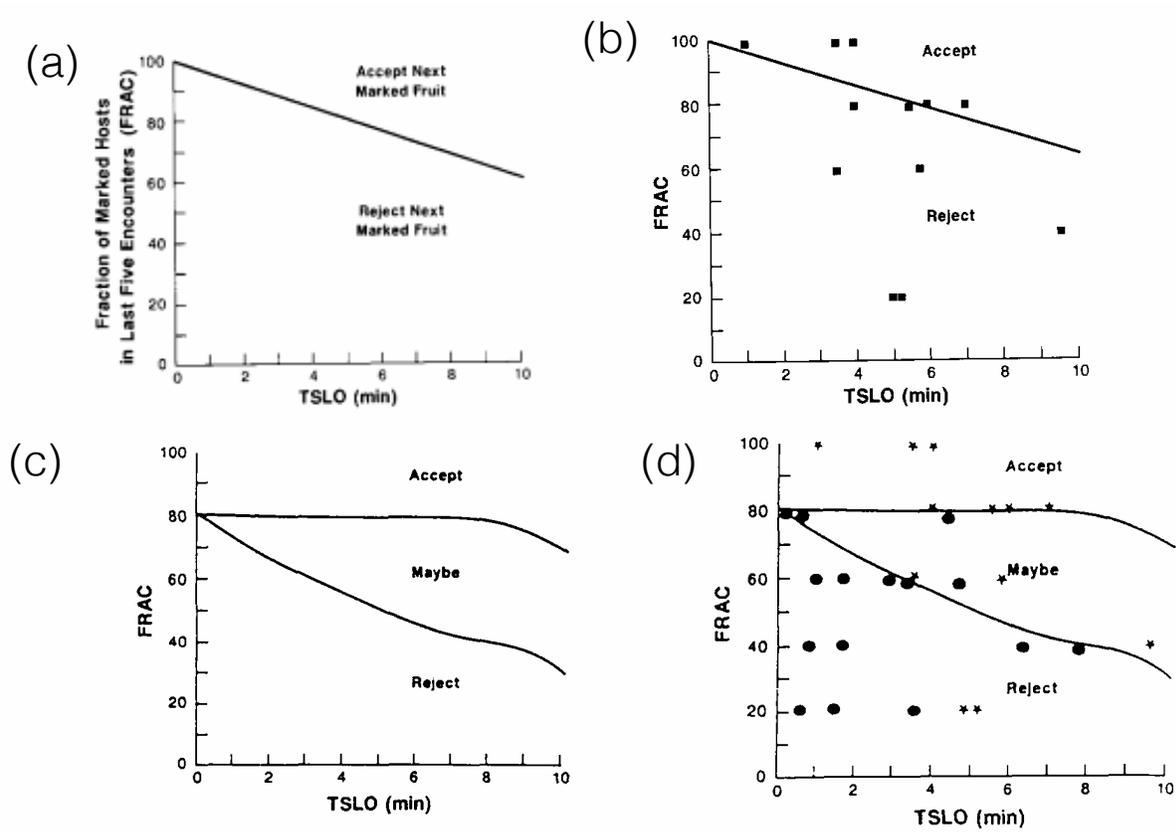


Figure 7: a) Theoretical predictions of host acceptance in the plane determined by the time since last oviposition (TSLO) and fraction of marked hosts in the last five encounters (FRAC). According to the simplest theory (Iwasa et al 1984) if a marked host is encountered for the combination of events above the line, the fly is predicted to oviposit in the marked host. If a marked host is encountered for points below the line, the fly is predicted to reject the marked host. b) Ovipositions by female *R. pomonella* in cages as a function of TSLO and FRAC. Approximately half of the points are in accordance with the theory and half do not agree with the theory. c) A simple informational approach divides the 'TSLO-FRAC' plane into regions in which we predict and individual will always oviposit, never oviposit, and sometimes (depending on details of the informational state) oviposit. This figure is drawn for the case in which three eggs remain. d) Comparison of observations (panel b) the theoretical curves (panel c). Instances in which a marked fruit was accepted for oviposition are indicated by a star. Also shown are fifteen quasi-randomly selected instances in which a marked fruit was rejected (circles). Note that nearly all acceptances fall in the 'accept' or 'maybe' regions and that the rejections fall in the 'maybe' and 'reject' regions. From Mangel and Roitberg (1989).

403 In a comparison of experimental and field studies of foraging by chacma baboons
404 *Papio ursine*, Marshall et al (2013) found greater support for Bayesian updating than the
405 classical Marginal Value Theorem in laboratory experiments but concluded that in the
406 field the animals likely incorporated much previous experience – as in this sliding memory
407 model.

408 **Classical Bayesian Approaches: The Conjugate Prior**

409 In the last quarter of the 20th century, Bayesian methods went through a revolution
410 because powerful desktop computers made computing accessible to virtually everyone.
411 But there was a lot of history between Bayes’s paper (Bayes 1763) and the Macintosh
412 computer (e.g. Isaacson 2014); I think that it is worthwhile to understand some of the
413 classical methods, particularly the method of the conjugate prior.

414 The mathematics of conjugate priors *seems* complicated but if you take it slowly, you
415 will see that it is actually not, because we take advantage of the fundamental property
416 that probabilities have to sum to 1 (i.e. something has to happen). That is, if $f(x)$
417 is any probability density and we sum or integrate over all of the values it can take,
418 $\sum f(x) = 1$ or $\int f(x)dx = 1$. Working through these examples carefully will help develop
419 your mathematical chops. And you may even find that they are fun to use!

420 I will do the first example very slowly, to be sure that there are no missing steps. After
421 doing that example, however, you will see how we can proceed more quickly with other
422 examples.

423 **The Poisson-Gamma Model**

To help ground our thinking, let’s consider the canonical equation for activity choice in
the following situation. Imagine a forager who can choose between one of two patches in
each time period. The increase in state when patch 1 is chosen is fixed and deterministic
 Y_1 but the increase when patch 2 is chosen is a random variable. In particular, let us
assume that the number of prey items encountered in 1 unit of search time follows a
Poisson process with parameter λ so that (Hilborn and Mangel 1997, Mangel 2006)

$$\Pr\{ n \text{ encounters in 1 time unit} \mid \lambda \} = \frac{\lambda^n}{n!} e^{-\lambda} \quad (46)$$

A quick note: Since when solving the SDP that will come along shortly, we work in single
periods of time, but λ is still a rate with units of 1/time. So the terms involving λ above

can be thought of as $\lambda \cdot 1$, where the ‘1’ has units of time. For a derivation of the Poisson process see Mangel (2006, pg 95-102).

Furthermore, let us assume that the the rate at which encounters occur is itself a random variable, which we denote by $\tilde{\lambda}$. Thus, when the individual forages in patch 2, encounters both lead to a change of state and provide information about the value of $\tilde{\lambda}$. For example, since the mean of a Poisson process with parameter λ is λ , so 2 encounters in one unit of time suggest a different value for the unknown encounter rate than 10 encounters in a unit of time. It is this issue that we need to address.

Rates are continuous variables that can take any positive value, so we specify a probability density $f(\lambda)$ for the unknown rate in the sense that

$$\Pr\{\lambda \leq \tilde{\lambda} \leq \lambda + \Delta\lambda\} = f(\lambda)\Delta\lambda + o(\Delta\lambda) \quad (47)$$

424 where $\Delta\lambda$ is a small increment in the parameter, and $o(\Delta\lambda)$ represents higher powers of
 425 $\Delta\lambda$, such as $\Delta\lambda^2, \Delta\lambda^3$, etc.

When $\Delta\lambda$ is small, the left-hand side of Eqn 47 is approximately the probability that $\tilde{\lambda}$ takes values close to λ , so taking liberty with the mathematical preciseness, I will sometimes write it as $\tilde{\lambda} \approx \lambda$. Since the rate must be positive but is otherwise not specified except that it must take some value, we have a normalization condition

$$\int_0^{\infty} f(\lambda)d\lambda = 1 \quad (48)$$

426 Note that the left hand side of Eqn 46 could be written as

427 $\Pr\{n \text{ encounters in 1 time unit} \mid \tilde{\lambda} = \lambda\}$.

We now ask: Given that n encounters occurred in 1 unit of time, what do we now know about $\tilde{\lambda}$? That is, we wish to know the probability that $\tilde{\lambda} \approx \lambda$ given the n encounters. Using Bayes’s theorem, we have

$$\Pr\{\tilde{\lambda} \approx \lambda \mid n \text{ encounters in 1 unit of time}\} = \frac{\Pr\{\tilde{\lambda} \approx \lambda \text{ and } n \text{ encounters in 1 unit of time}\}}{\Pr\{n \text{ encounters in 1 unit of time}\}} \quad (49)$$

Let us unpack the numerator and denominator in Eqn 49. In light of the fundamental

rules of probability

$$\begin{aligned} & \Pr\{\tilde{\lambda} \approx \lambda \text{ and } n \text{ encounters in 1 unit of time}\} = \\ & \Pr\{n \text{ encounters in 1 unit of time} | \tilde{\lambda} \approx \lambda\} \cdot \Pr\{\tilde{\lambda} \approx \lambda\} \end{aligned} \quad (50)$$

The first term on the right hand side of Eqn 50 is the Poisson distribution (Eqn 46) and the second term on the right hand side is $f(\lambda)$ (Eqn 47), which we have yet to specify. Hence we conclude that

$$\Pr\{\tilde{\lambda} \approx \lambda \text{ and } n \text{ encounters in 1 unit of time}\} = \frac{e^{-\lambda} \lambda^n}{n!} f(\lambda) \quad (51)$$

The denominator in Eqn 49 is the probability of obtaining n encounters, taken across all possible values of the rate of encounter λ so that

$$\Pr\{n \text{ encounters in 1 unit of time}\} = \int_0^\infty \frac{e^{-\lambda} \lambda^n}{n!} f(\lambda) d\lambda \quad (52)$$

We call $f(\lambda)$ the prior density for the unknown rate of encounter and the left hand side of Eqn 51 the posterior distribution of the rate of encounter, given that n encounters occurred in 1 unit of time. We denote the posterior density by $f(\lambda|n)$, always keeping in mind the 1 unit of time. Using Eqns 51 and 52 and noting that the $n!$'s cancel in numerator and denominator, we obtain

$$f(\lambda|n) = \frac{e^{-\lambda} \lambda^n f(\lambda)}{\int_0^\infty [e^{-\lambda} \lambda^n f(\lambda)] d\lambda} \quad (53)$$

428 How does this help us? Well, to begin we recognize that the denominator – since we
 429 are integrating over λ - is simply a constant. It would be better mathematical grammar
 430 to write it as $\int_0^\infty [e^{-\lambda'} (\lambda')^n f(\lambda')] d\lambda'$ to remind ourselves that the encounter rates in the
 431 numerator and denominator are different (this is good mathematical grammar). But we
 432 still have not chosen $f(\lambda)$.

433 Happily for us, there is a special function of applied mathematics that helps. The
 434 phrase “special function” should not freak you out; indeed you are already familiar – from
 435 elementary calculus and population biology – with a number of special functions (although
 436 we don't tend to call them that; sometimes they are called the elementary transcendental
 437 functions; see Abramowitz and Stegun (1965)): e^x , $\log(x)$, $\sin(x)$, $\cos(x)$, $\sin^{-1}(x)$ and
 438 $\cos^{-1}(x)$ [the exponential logarithm, sine, cosine, inverse sine and inverse cosine functions

439 respectively]. Indeed, if you have worked in population biology, you have learned to love
 440 at least the first four of these functions. Think about their characteristics: they arise in
 441 many situations, are the solutions of certain differential equations, and have been well
 442 tabulated (in the old days, that meant in books, as in Abramowitz and Stegun (1965);
 443 these days it means built into computer software). There is no reason that we cannot
 444 continue to add include other special functions in our study of biology. For example,
 445 Lehtonen (2016) argues that the Lambert W function should also be in the toolbox of
 446 population biologists.

The gamma function is a slightly less elementary transcendental function, but as you will see, it is very useful. Let's write the density $f(\lambda)$ on the right hand side of Eqn 53 as

$$f(\lambda|\alpha, \nu) = \frac{\alpha^\nu}{\Gamma(\nu)} e^{-\alpha\lambda} \lambda^{\nu-1} \quad (54)$$

447 where I use the vertical bar to denote that whereas λ is the variable, α (“alpha”) and ν
 448 (“new”) are parameters. On the right hand side of this equation $\Gamma(\nu)$ (“gamma of new”) is the gamma function of ν . I explain the gamma function in some detail in Mangel (2006, pg 104ff). For now, it is sufficient to know that $\Gamma(1) = 1$ and for any other number $\Gamma(\nu + 1) = \nu\Gamma(\nu)$. You should convince yourself that if ν is an integer, then $\Gamma(\nu) = 1 \cdot 2 \cdot 3 \cdots (\nu - 1) = (\nu - 1)!$ so that you can think of the gamma function as a generalization of the factorial function.

449 Furthermore, since α and ν are parameters, we could just as easily have written
 450 $f(\lambda|\alpha, \nu) = c_{norm} e^{-\alpha\lambda} \lambda^{\nu-1}$ where c_{norm} is a normalization constant, ensuring that the
 451 integral of $f(\lambda|\alpha, \nu)$ over all possible encounter rates is 1. This observation will become
 452 helpful shortly. Remembering that one cannot exponentiate or use as a power numbers
 453 that have units, since λ is a rate, with units 1/time, if the the product $\alpha\lambda$ has no units
 454 it must be that α has units of time. Since we use ν as an exponent, it must be a pure
 455 number.

456 The mean and coefficient of variation of $\tilde{\lambda}$ are $E(\tilde{\lambda}) = \frac{\nu}{\alpha}$ and $CV(\tilde{\lambda}) = \frac{1}{\sqrt{\nu}}$ (Hilborn
 457 and Mangel 1997; Mangel 2006). In Figure 8, I show three gamma densities, illustrating
 458 why ν is called the shape parameter.

459 The reason for choosing the gamma density is that the Poisson distribution of encounters Eqn 46 involves an exponential of λ and a power of λ . The gamma density Eqn 54 involves an exponential of λ and a power of λ . Since the denominator in Eqn 53 is a constant, we conclude that if the prior is a gamma density with parameters α and ν and

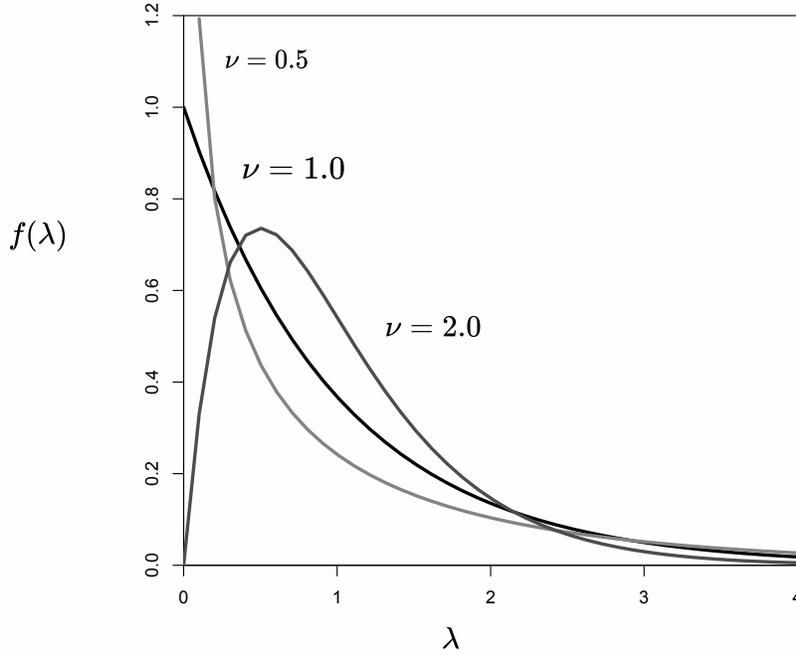


Figure 8: Three gamma densities with the same mean but different values of the dispersion parameter ν , showing the range of shapes that the gamma density can have.

468 the forager has n encounters in 1 unit of time then the posterior density will also be a
 469 gamma.

Since $f(\lambda|\alpha, \nu)$ is a probability density, we know that $\int_0^\infty f(\lambda|\alpha, \nu)d\lambda = 1$ which means that

$$\int_0^\infty e^{-\alpha\lambda}\lambda^{\nu-1}d\lambda = \frac{\Gamma(\nu)}{\alpha^\nu} \quad (55)$$

470 You can think of this as the definition of the gamma function $\Gamma(\nu)$.

When we substitute Eqn 54 into Eqn 53 and take account that $\frac{\alpha^\nu}{\Gamma(\nu)}$ appears in numerator and denominator so can be cancelled out, we obtain

$$f(\lambda|n, \alpha, \nu) = \frac{e^{-\lambda}\lambda^n e^{-\alpha\lambda}\lambda^{\nu-1}}{\int_0^\infty [e^{-\lambda}\lambda^n e^{-\alpha\lambda}\lambda^{\nu-1}]d\lambda} \quad (56)$$

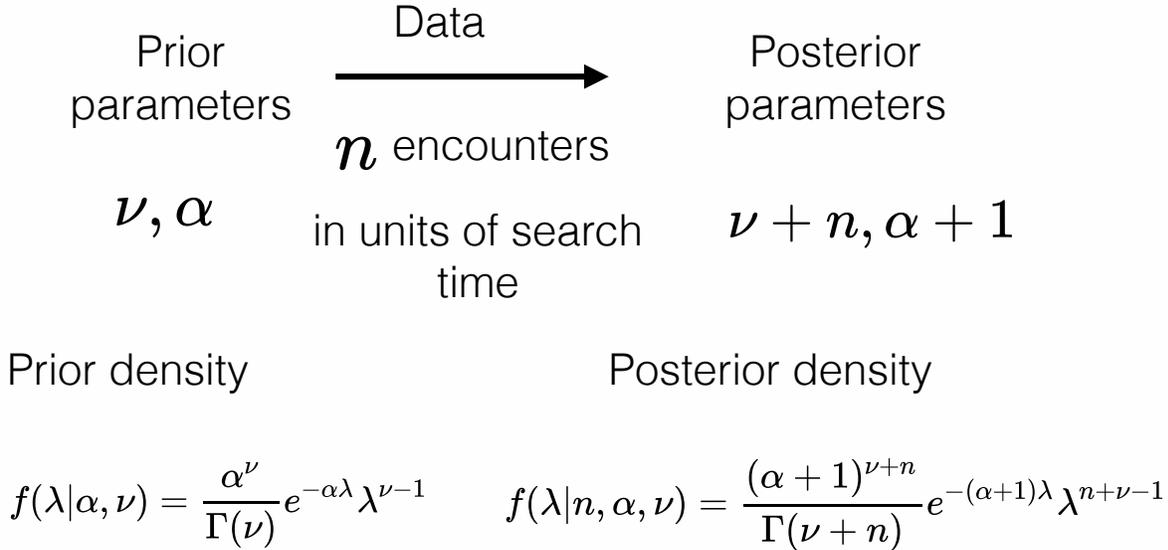


Figure 9: Schematic illustration of Bayesian updating for the Gamma-Poisson model.

Combining terms we obtain

$$f(\lambda|n, \alpha, \nu) = \frac{e^{-(\alpha+1)\lambda} \lambda^{n+\nu-1}}{\int_0^\infty [e^{-(\alpha+1)\lambda} \lambda^{n+\nu-1}] d\lambda} \quad (57)$$

Comparing the numerators of Eqns 54 and 57, we see that instead of α we have $\alpha + 1$ and instead of ν we have $\nu + n$. Thus the denominator in Eqn 57 must be $\frac{\Gamma(\nu+n)}{(\alpha+1)^{\nu+n}}$ so that the posterior density is

$$f(\lambda|n, \alpha, \nu) = \frac{(\alpha + 1)^{\nu+n}}{\Gamma(\nu + n)} e^{-(\alpha+1)\lambda} \lambda^{n+\nu-1} \quad (58)$$

471 We have updated the mean value of $\tilde{\lambda}$ from $\frac{\nu}{\alpha}$ to $\frac{\nu+n}{\alpha+1}$ and the coefficient of variation from
 472 $\frac{1}{\sqrt{\nu}}$ to $\frac{1}{\sqrt{\nu+n}}$. The gamma density with these new parameters serves as the prior in the
 473 next period (Figure 9). The key observation is that we begin with a gamma density for
 474 $\tilde{\lambda}$, collect the information of n encounters in 1 unit time and end up with another gamma
 475 density, with changed (updated) parameters, because we picked a form for $f(\lambda)$ that has
 476 a structure similar to the Poisson process.

The gamma density is often written as

$$f(\lambda|\beta, \nu) = \frac{1}{\beta^\nu \Gamma(\nu)} e^{-\lambda/\beta} \lambda^{\nu-1} \quad (59)$$

477 where $\beta = 1/\alpha$ is a rate like λ . This form of the gamma density is used in both R and
 478 MATLAB. β (or α) is called the scale parameter. Note that if $\nu = 1$, then since $\Gamma(1) = 1$,
 479 Eqn 54 becomes, $f(\lambda|\alpha, 1) = \alpha e^{-\alpha\lambda}$, which is the exponential density.

480 **The SDP Equation with the Poisson-Gamma model**

We are now ready to turn to the SDP. To focus on information, we keep the state dynamics as simple as possible, letting $X(s)$ denote energy reserves at the start of period S and assume that the cost of foraging in each patch is the same, denoted by c . Then if the individual chooses patch 1, with certainty we know

$$X(s+1) = X(s) - c + Y_1 \quad (60)$$

But encounters in patch 2 are random. If each prey item encountered provides energy ϵ , when n prey items are encountered in a unit of time

$$X(s+1) = X(s) - c + \epsilon n \quad (61)$$

Let us use $p_n(\alpha, \nu)$ to denote the probability of encountering n prey items in a unit of time when the informational parameters are α and ν . It is the right hand side of Eqn 52 with $f(\lambda)$ given by the gamma density so that

$$p_n(\alpha, \nu) = \int_0^\infty \frac{e^{-\lambda} \lambda^n}{n!} \cdot \frac{\alpha^\nu}{\Gamma(\nu)} e^{-\alpha\lambda} \lambda^{\nu-1} = \frac{\alpha^\nu}{n! \Gamma(\nu)} \int_0^\infty e^{-(\alpha+1)\lambda} \lambda^{\nu+n-1} d\lambda \quad (62)$$

We have already discussed the integral in Eqn 62, between Eqns 57 and 58; it is $\frac{\Gamma(\nu+n)}{(\alpha+1)^{\nu+n}}$ so we conclude

$$p_n(\alpha, \nu) = \frac{\alpha^\nu}{n! \Gamma(\nu)} \frac{\Gamma(\nu+n)}{(\alpha+1)^{\nu+n}} = \frac{1}{n!} \frac{\alpha^\nu}{(\alpha+1)^{\nu+n}} \frac{\Gamma(\nu+n)}{\Gamma(\nu)} \quad (63)$$

481 Eqn 63 is called the negative binomial distribution. In this form or slight variants it
 482 has a long and rich history in ecology (see Mangel 2006, pg 107-112, where you can also
 483 get some hints about computing the distribution and a few exercises to help you build

484 intuition).

Since information, captured in α and ν is now a state variable, the fitness function for this version of the canonical equation of activity choice is $F(x, \alpha, \nu, s)$ defined as

$$F(x, \alpha, \nu, s) = \text{maximum expected value of } \Phi(X(S)) \text{ given that } X(s) = x \\ \text{and current informational parameters are } \alpha \text{ and } \nu \quad (64)$$

If $V_i(x, \alpha, \nu, s)$ is the fitness value of foraging in patch i when $X(s) = x$ and the informational parameters are α and ν then

$$F(x, \alpha, \nu, s) = \max[V_1(x, \alpha, \nu, s), V_2(x, \alpha, \nu, s)] \quad (65)$$

When the individual forages in patch 1, its state is incremented in a deterministic fashion (Eqn 60) and the informational states will not change so that if the rate of mortality in patch 1 is m_1 we have

$$V_1(x, \alpha, \nu, s) = e^{-m_1} F(x - c + Y_1, \alpha, \nu) \quad (66)$$

On the other hand, when the forager visits patch 2, the best that we can say is that $n = 0, 1, 2, \dots$ prey items may be encountered, according to the negative binomial distribution in Eqn 63. When n prey items are encountered in a single period of search, the physiological state is changing according to Eqn 61 and the informational parameters are updated to $\alpha + 1$ and $\nu + n$ so that

$$V_1(x, \alpha, \nu, s) = \mathcal{E}_n \{ e^{-m_1} F(x - c + \epsilon n, \alpha, \nu) \} \\ = \sum_n p_n(\alpha, \nu) \{ e^{-m_1} F(x - c + \epsilon n, \alpha + 1, \nu + n) \} \quad (67)$$

485 Eqn 67 raises the challenges of informational problems. First, in principle the sum over
 486 n in Eqn 67 is unbounded. We can deal with that by finding a value n_{max} for which the
 487 sum $\sum_{n=0}^{n_{max}} p_n(\alpha, \nu)$ is sufficiently close to 1 (say 0.999 or 0.9999) and use a renormalized
 488 version of $p_n(\alpha, \nu)$, in which n ranges from 0 to n_{max} . The second difficulty comes when
 489 setting memory for the fitness functions. Clearly the maximum value that x can take is
 490 x_{max} . If the individual were to visit patch 2 in every period, then the maximum value that
 491 α could take is its starting value plus S . The bigger problem is the maximum value that
 492 ν can take. In principle at least, in every period the individual can encounter n_{max} prey

493 items, so that again in principle ν could be as large as its starting value plus $n_{max} \cdot S$.
 494 This could be a large number, and this issue is called the “curse of dimensionality”
 495 in dynamic programming. Over the years various methods around this problem have
 496 been developed from simple myopic approaches (Mangel and Clark 1983) to approximate
 497 dynamic programming (Powell 2011, 2014, 2015ab, Powell et al 2012).

498 **The Beta-Binomial Model**

499 In the Poisson-gamma model, we allowed an arbitrary number of events in a single unit of
 500 time. However, in many situations tin a single unit of time either something happens (e.g.
 501 a parasitoid encounters a host) or nothing happens (e.g. no host encountered). Then after
 502 N units time, the (random) number of events will range over $n = 0, 1, 2, \dots, N$, and the
 503 simplest probability distribution characterizing the process is the binomial distribution
 504 with (unknown) probability \tilde{p} of an event in a unit time. We will now explore how one
 505 can learn about \tilde{p} with Bayesian methods.

We proceed in manner analogous to the Poisson and gamma. Conditioned on $\tilde{p} = p$, the probability of n events in N periods is given by the binomial distribution

$$\Pr\{n \text{ events in } N \text{ periods} | \tilde{p} = p\} = \binom{N}{n} p^n (1-p)^{N-n} \quad (68)$$

We ask: beginning with a prior for \tilde{p} and the data of n events in N periods, what can we say about the posterior for \tilde{p} ? The analogy to Eqn 49 is

$$\Pr\{\tilde{p} \approx p | n \text{ encounters in } N \text{ periods of search}\} = \frac{\Pr\{\tilde{p} \approx p \text{ and } n \text{ encounters in } N \text{ periods}\}}{\Pr\{n \text{ encounters in } N \text{ periods}\}} \quad (69)$$

Following the same logic that lead to Eqn 53, if $f(p)$ denotes the prior probability density of \tilde{p} and $f(p|n, N)$ the posterior after N periods with n encounters is

$$f(p|n, N) = \frac{p^n (1-p)^{N-n} f(p)}{\int_0^1 [p^n (1-p)^{N-n} f(p)] dp} \quad (70)$$

How do we choose the prior $f(p)$? We picked the gamma density, because it had the same structure as probability of the data. This suggests that we make the choice

$$f(p) = c_{norm} p^{\alpha-1} (1-p)^{\beta-1} \quad (71)$$

where c_{norm} is a constant chosen so that

$$\int_0^1 c_{norm} p^{\alpha-1} (1-p)^{\beta-1} dp = 1 \quad (72)$$

506 Since c_{norm} is a constant, it comes out of the integral so that $c_{norm} = \frac{1}{\int_0^1 p^{\alpha-1} (1-p)^{\beta-1} dp}$.

Once more classical applied mathematics comes to our aid: the integral in Eqn 72 has been studied for many years. It is called the beta function

$$\int_0^1 p^{\alpha-1} (1-p)^{\beta-1} dp = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)} = \mathcal{B}(\alpha, \beta) \quad (73)$$

507 where our new friend the gamma function appears!

Thus the prior density is

$$f(p|\alpha, \beta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} p^{\alpha-1} (1-p)^{\beta-1} \quad (74)$$

508 This density can take a wide variety of shapes (Figure 10) For example, if $\alpha = \beta = 1$ then
 509 both exponents are 0 and the density is uniform on 0 to 1; this would be the situation in
 510 which every value of \tilde{p} is equally likely. Exactly what that means in terms of knowledge
 511 that the animal is an interesting question. Sometimes people want that a uniform prior
 512 is “uninformative”, but that is not a good description since asserting that every value of
 513 p is equally likely actually requires knowing a lot about the world. There is a specific
 514 definition for what makes a prior non-informative (roughly that the data change the
 515 location but not the shape of the posterior; see Martz and Waller (1982) for an accessible
 516 introduction and Mangel and Beder (1982) for comparison of uniform and non-informative
 517 priors in a search problem in fisheries).³

If α and β are the same, then the density is symmetric around $p = 0.5$. When both of the parameters are greater than 1, the density is 0 at both $p = 0$ and $p = 1$ and when both parameters are less than 1, the density rises to infinity as $p \rightarrow 0, 1$. But many

³One important difference between the beta and gamma densities, is that p can only take values between 0 and 1 while λ can take all values greater than 0 – although we expect that extremely high encounter rates are somewhere between improbable and impossible. This becomes clear if we set $\alpha = 0$ and $\nu = 1$ in Eqn 54. Then if we ignore the normalization constant we have $f(\lambda|0, 1) = 1$ and we have to ignore the normalization constant because $\int_0^\infty 1 \cdot \lambda d\lambda$ is infinite. A prior such as this, which cannot be normalized, is called an improper prior. But note that if we began with such a prior and had n encounters in 1 unit of time, the posterior is a gamma density with parameters $\alpha = 1$ and $\nu = n$ so that the posterior is perfectly proper. Another way around this problem, is to assume that λ is bounded above by a value λ_{max} so that the prior would be $f(\lambda) = \frac{1}{\lambda_{max}}$, which clearly integrates between 0 and λ_{max} to 1. We will use this prior in the next section.

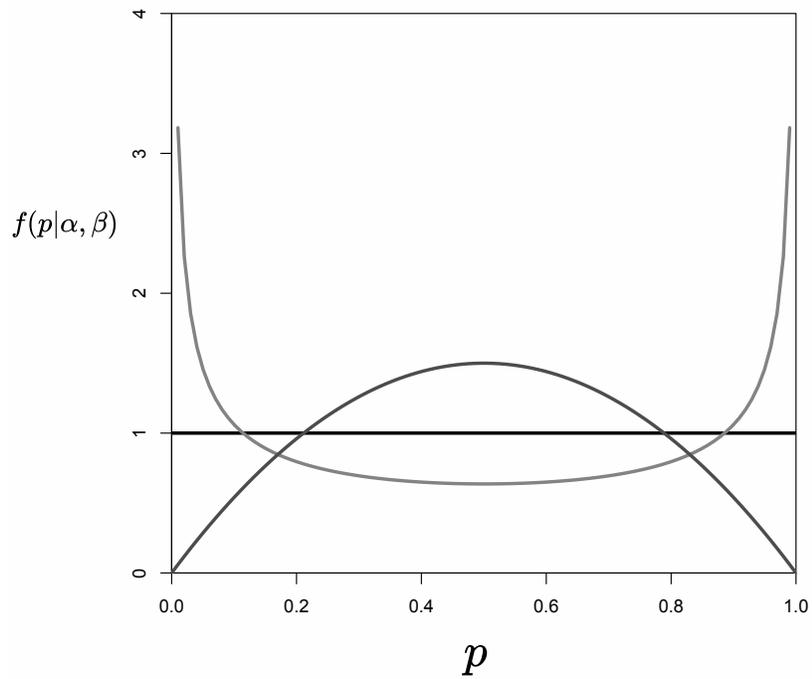


Figure 10: Three beta densities with the same mean but different values of the parameters α and β , which I have set equal to each other. In the flat line, the parameters are 1, in the curve opening upwards they are $1/2$, in the curve going to 0 at $p = 0$ and $p = 1$ they are 2.. To see that you understand the analysis we're doing, look at Eqn 74 and then explain the shapes of these curves.

other possibilities exist and I encourage you to play around with the density to get some intuition about it (the beta density is a built in function in R, so it is easy to explore).

If we substitute Eqn 74 into Eqn 70 and take account that the terms involving the gamma functions cancel (since the gamma functions are constants when we integrate over p), the posterior density is

$$f(p|\alpha, \beta, n, N) = \frac{p^n(1-p)^{N-n}p^{\alpha-1}(1-p)^{\beta-1}}{\int_0^{\infty} p^n(1-p)^{N-n}p^{\alpha-1}(1-p)^{\beta-1}dp} \quad (75)$$

The numerator in this equation simplifies to $p^{n+\alpha-1}(1-p)^{N-n+\beta-1}$. The denominator is the integral of the numerator. Comparing with Eqn 74, we conclude that the posterior density is

$$f(p|\alpha, \beta, n, N) = \frac{\Gamma(\alpha + \beta + N)}{\Gamma(\alpha + n)\Gamma(\beta + N - n)} p^{n+\alpha-1}(1-p)^{N-n+\beta-1} \quad (76)$$

518 In other words, if the individual started the day with a prior beta density with parameters
 519 α and β characterizing uncertainty \tilde{p} and had n encounters in N periods of search, then
 520 the posterior beta density characterizing the uncertainty in \tilde{p} would have parameters $n + \alpha$
 521 and $N - n + \beta$.

522 Let us focus now on a single period for which the current values of the parameters
 523 are α and β . If an encounter occurs during this period, the updated parameters at the
 524 start of the next period are $\alpha + 1$ and β ; if an encounter does not occur, the updated
 525 parameters are α and $\beta + 1$.

If the unknown probability of encounter \tilde{p} has a beta density with parameters α and β then its mean and variance are (Hilborn and Mangel 1997, Mangel 2006)

$$E[\tilde{p}] = \frac{\alpha}{\alpha + \beta}$$

$$Var[\tilde{p}] = \frac{\alpha\beta}{(\alpha + \beta)^2(\alpha + \beta + 1)} \quad (77)$$

526 which can be verified or computing the first and second moments of \tilde{p} with the beta
 527 density Eqn 74 and then simplifying using properties of the gamma function.

We will often want to specify a mean m and variance v for \tilde{p} and from them determine

the parameters. As long as $v < m(1 - m)$ the parameters are

$$\begin{aligned}\alpha &= m \left(\frac{m(1 - m)}{v} - 1 \right) \\ \beta &= (1 - m) \left(\frac{m(1 - m)}{v} - 1 \right)\end{aligned}\tag{78}$$

528 You can verify that this is true by using the values in Eqn 78 in Eqn 77 and seeing that
529 the mean m and variance v emerge.

530 **The SDP Equation with the Beta-Binomial Model**

The following two models in arose in discussions with Sasha Dall when we both were visiting the University of Bergen in April 2016. Sasha has thought for many years about information and its fitness consequences, and in 2010 edited a special section of *Oikos* on biological information in an ecological context (volume 119, pgs 201-316). Let's consider the canonical equation for resource allocation with a fixed end time, but using p for the probability of encountering a host in a single time period, so that we can save Greek letters for updating, and make the dependence on the probability of encountering a host and mortality explicit. Then we still have the end condition $F(x, S|p, m) = 0$ and the SDP (Mangel 2015,

$$\begin{aligned}F(x, s|p, m) &= e^{-m}(1 - p)F(x, s + 1|p, m) \\ &\quad + p \max_{c \leq x} [f(c) + e^{-m(c)} F(x - c, s + \tau(c)|p, m)]\end{aligned}\tag{79}$$

531 If we knew their specific values, say \bar{p} and \bar{m} then solving Eqn 79 is straightforward and
532 the measure of lifetime fitness is $F(x_{max}, 1|\bar{p}, \bar{m})$.

533 When the probability of encounter is unknown (for simplicity, I am going to keep the
534 rate of natural mortality fixed) and the parasitoid learns about the world a natural way
535 of characterizing the uncertainty about p is with a beta density $\mathcal{B}(p|\alpha, \beta)$ as in Eqn 74.
536 We can imagine that the prior values of α and β are set by a chemical cue. In each period
537 the parasitoid either encounters a host, in which case α increases by 1 and β stays the
538 same, or the parasitoid does not encounter a host, in which case α stays the same and β
539 increases by 1.

Writing fitness as $F(x, s|\alpha, \beta, m)$ the SDP equation becomes

$$F(x, s|\alpha, \beta, m) = \mathcal{E}_p\{e^{-m}(1-p)F(x, s+1|\alpha, \beta+1, m) + p \max_{c \leq x} [f(c) + e^{-m(c)}F(x-c, s+\tau(c)|\alpha+1, \beta, m)]\} \quad (80)$$

540 The \mathcal{E}_p indicates an expectation over p . In practice, when the parameters of the beta
 541 density are less than 1, we can discretize p so that its minimum is a very small value,
 542 say 0.001, and its maximum value equally close to 1, say 0.999 and then renormalize the
 543 probability density. Each value of p is given probability according to Eqn 74. Given the
 544 value of p , we do the usual SDP actions contained within the $\{\dots\}$. Since the informational
 545 states change by at most 1 each period, they each require a dimension of $S+1$.

546 We can determine the fitness value of learning as follows. Letting α_0 and β_0 de-
 547 note the initial values of the parameters, parasitoid that learns has first period fitness
 548 $F(x_{max}, 1, |\alpha_0, \beta_0, m)$, determined by the solution of Eqn 80. A parasitoid that does not
 549 learn would have first period fitness determined by Eqn 70 with p always equal to $\frac{\alpha_0}{\alpha_0+\beta_0}$.

550 **Other Conjugate Priors: Threshold Norms of Reaction Emerge From Bayesian** 551 **Analysis**

There are many other conjugate priors. For example, suppose that the data \tilde{X} come from a normal distribution with unknown mean $\tilde{\mu}$ and standard deviation $\sigma = 1$ so that

$$\Pr\{\tilde{X} \approx x | \tilde{\mu} = \mu\} = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2}\right) \quad (81)$$

552 In Appendix 2, I show that a good choice for the conjugate prior for $\tilde{\mu}$ is another normal
 553 distribution and show how to do the Bayesian updating. Tachiki and Koizumi (2016)
 554 used Bayesian analysis and the normal-normal probability and prior to model the norm
 555 of reaction in Masu salmon *Oncorhynchus masou* (Figure 11).

556 **Modern Bayesian Methods**

557 If you like the kinds of calculations that we have done, check out one of the many wonderful
 558 books on Bayesian statistical methods (e.g Gelman et al 2009, Dorazio 2016, McElreath
 559 2016). The approach of the conjugate prior was invented when computing was hard to
 560 do. It has the advantage of requiring only simple summaries of the data. But we now
 561 live in a world in which computing is much easier and modern computational methods

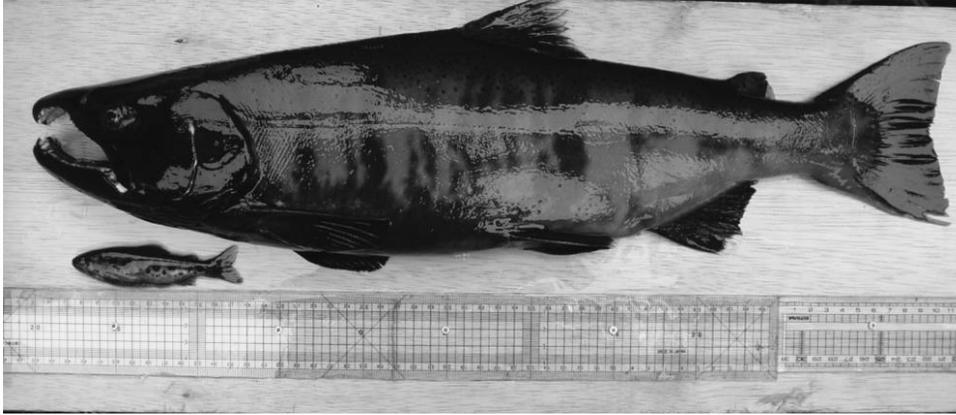


Figure 11: Migrant (above) and resident (below) forms of male masu salmon (*Oncorhynchus masou*). Individuals must choose one of these tactics while in the juvenile stage. Smaller juveniles tend to migrate to the ocean for feeding and then return to their natal river, whereas larger, more dominant juveniles tend to stay in the natal river. Tachiki and Koizumi (2016), from whom this image is taken, show how threshold norms of reaction emerge from Bayesian analysis.

562 have lead to a raft of Bayesian methods that are computer intensive; Dorazio (2016) and
563 McElreath (2016) are terrific introductions to such methods for biologists. For example,
564 in the Poisson-gamma case, we might have evidence that the prior is something other
565 than a gamma density. If we knew that the second patch is either very good (high values
566 of $\tilde{\lambda}$ or very bad (low values), the gamma density cannot capture this feature. The beta
567 density could be used, but if we were to use a a beta density as the prior (scaled to the
568 maximum value of the encounter rate), we'd lose the advantage of the having the same
569 form as the encounter process (trying writing out the combination of a beta density prior
570 and a Poisson process).

571 Alternatively, we may not know the parameters of the gamma density in the Poisson-
572 gamma case accurately and would like to give them a distribution characterized by its
573 own parameters. That is, we would start not with single values of α and ν but a range of
574 values and learn about the hyper-parameters that characterize this range.

575 A variety of modern tools, based on one way or another on Monte Carlo simulation,
576 exist for situations like these and many good books explain how this is done (e.g. Gelman
577 et al 2009, McElreath 2016). For an example of the application of such methods for a
578 fishery management problem using SDP, check out Boettiger et al (2015).

579 Posteriors in a Changing World

580 Environmental change will have an effect on how we interpret the posterior distribution
581 (Lloyd and Leslie 2013). In particular, while an individual learns about the environment,
582 there is also the possibility that the environment changes, either in time or space.

Let us consider changes in time first. For the Poisson-gamma updating model, let $\nu(s)$ and $\alpha(s)$ denote the values of the parameters at time s and $n(s)$ the resources encountered in the time interval between s and $s + 1$ so that the updating rules can be written as

$$\begin{aligned}\alpha(s + 1) &= \alpha(s) + 1 \\ \nu(s + 1) &= \nu(s) + n(s)\end{aligned}\tag{82}$$

583 To use these equations, we need to provide initial conditions: $\nu(0) = \nu_0$ and $\alpha(0) = \alpha_0$.
584 Clearly $\alpha(s) = \alpha_0 + s$, but $\nu(s)$ depends upon the data from encounters.

Now let us suppose that there is probability γ that the environment reverts to the initial conditions. Then we replace Eqn 82 by

$$\begin{aligned}\alpha(s + 1) &= (1 - \gamma)[\alpha(s) + 1] + \gamma\alpha_0 \\ \nu(s + 1) &= (1 - \gamma)[\nu(s) + n(s)] + \gamma\nu_0\end{aligned}\tag{83}$$

585 The parameter γ can also be interpreted as the rate of forgetting (Mangel 1990): when
586 $\gamma = 0$ all previous information is remembered by the organism and when $\gamma = 1$, it
587 is Groundhog Day in which every period starts anew with parameters ν_0 and α_0 . For
588 intermediate values of γ , the individual forgets the past but at different rates; we can
589 even ask about the optimal rate of forgetting (Mangel 1990).

590 There is another way of thinking about Eqn 83, suggested to me by Judy Stamps: a
591 good learner is an individual who mainly relies current information to update (i.e. γ is
592 small) while a poor learner is more hesitant to give up past information (i.e. γ is close to
593 1). Thus, we can think of $\omega = 1 - \gamma$ as the ability to learn in an uncertain environment.

594 Parameters could also change in space. One of the triggers for the development of
595 behavioral ecology was the Marginal Value Theorem (MVT), which concerned organisms
596 foraging in patchy worlds (see Mangel (2006), pg 5-8 for a brief introduction). In general,
597 uses of the MVT assume that either all patches are identical or are drawn from a small
598 number of different kinds of patches with known properties.

To add an informational component, imagine that individual is born into a patch and

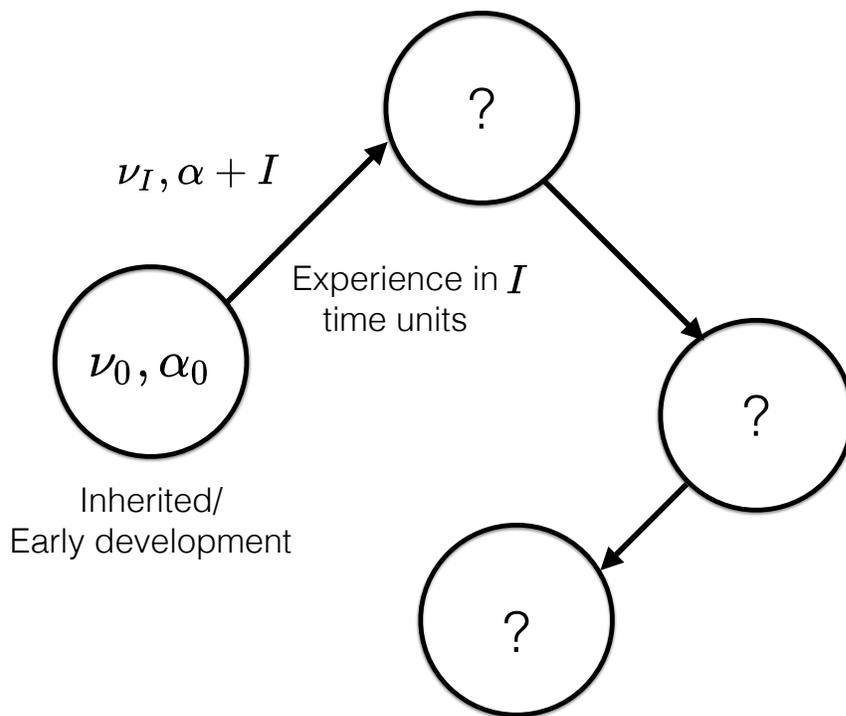


Figure 12: Schematic of the challenge of blending information that is inherited or experienced in early development with subsequent experience in which the patches may be different.

either inherits or obtains during early development assigns parameters ν_0, α_0 to this patch. After I periods of time, she leaves the natal patch with updated parameters ν_I, α_I and enters a new patch (Figure 12), where she gathers information about patch quality while foraging and then moves on to another patch. What prior should be associated with the next patch? If all patches are identical, then it would make sense to use the posterior from the natal patch and if all patches are independently uncertain, then it would make sense to use the prior from the natal patch. When the natural world sits somewhere in between those two situations, we weight the prior and posterior and set the parameters for the next patch visited to be

$$\begin{aligned}\nu_2 &= (1 - \omega)\nu_0 + \omega\nu_I \\ \alpha_2 &= (1 - \omega)\alpha_0 + \omega\alpha_I\end{aligned}\tag{84}$$

599 where ω is a measure of how independent the patches are. When $\omega = 0$ every new
600 patch encountered is approached ignoring prior experience; when $\omega = 1$ every new patch
601 encountered is approached using all of the prior experience.

602 In another way of thinking about this question, we do not weight prior and posterior
603 parameters but rather weight prior and posterior distributions. That is, we imagine that
604 the individual began with a prior for the encounter rate that is a uniform distribution
605 for λ between 0 and λ_{max} ; we use denote that by $f_u(\lambda|\lambda_{max}) = \frac{1}{\lambda_{max}}$. As we discussed
606 above, after I units of time in which n_e encounters occurred the posterior is the gamma
607 density $f(\lambda|\lambda_{max}, I, n_e)$. Instead of weighting the parameters as in Eqn 84, we weight
608 the two posterior distributions so that going into the new patch the distribution on the
609 encounter rate is $(1 - \omega)f_u(\lambda|\lambda_{max}) + \omega f(\lambda|\lambda_{max}, I, n_e)$. The method of the conjugate
610 prior is ill-suited for such a problem, but numerical methods can easily deal with it. Here
611 is a snippet of code, in the spirit of McElreath (2016) for implementing these ideas when
612 there are 8 encounters in two periods of developmental time.

613

A Snippet of Code for Posteriors in A Changing World

```
quartz()
N_p=100 # grid of 100 values of encounter rate lambda
lambda_max=20 #maximum value of lambda
lambda <- seq( from=0 , to=lambda_max , length=N_p )
f_u<- rep( 1 , N_p )/N_p #uniform prior on lambda
```

```

# -- think about why the denominator is
# N_p, the size of the grid, and not lambda_max
# as in the text

I=2
n_enc=8 #number of encounters in I periods of time

#We want to look at different values of omega
# -- the probability that
#the next patch is like this one
omega=seq(from=0, to=0.95,length=6)
plot(lambda, f_u, type="l", lwd=4, ylim=c(0,0.04),
ylab="Prior and Posterior")
for(i in 1:6)
{
  likelihood <- dpois(n_enc , lambda*I )
  f_p<- likelihood * f_0
  f_p <- f_p / sum(f_p)
  posterior = (1-omega[i])*f_u + omega[i]*f_p
  lines(lambda,posterior, type="l", lwd=4, col =i)
}
legend(15.,0.04, legend=c("omega=0.000", "=0.19",
"=0.38", "=0.57", "=0.76", "0.95"),
      col=c(1,2,3,4,5,6), lty=1, cex=0.6)
}

```

614 In Figure 13, I show the prior and the posteriors for six values of ω , including $\omega = 0$,
615 which should (and does give us the prior – the flat line).

616 Populations: Entropy and Information

With populations, there is a different way to approach information using the concept of entropy (Cover and Thomas 2006). For a discrete random variable X with probability distribution $p(x)$, so that $\Pr\{X = x\} = f(x)$, the entropy of the distribution is

$$H(X) = - \sum_x f(x) \log(f(x)) \quad (85)$$

Posterior

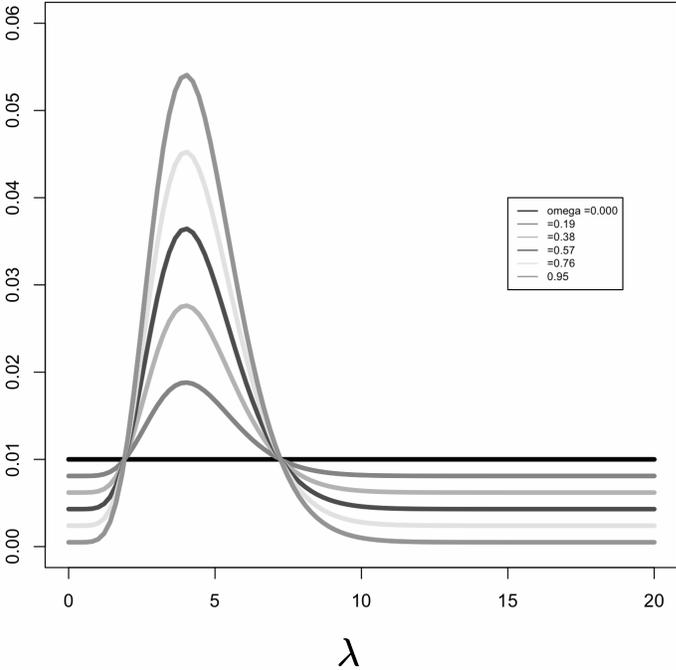


Figure 13: Output of the Rscript for posteriors in a changing world in which the individual collects information of 8 events in two periods of development and blends a uniform prior and the posterior from the gamma-Poisson conjugate Bayesian model to determine its operational prior. Here I show six posteriors and the prior, corresponding to different values of the the blending parameter ω . When $\omega = 0$ information is ignored and the posterior is the same as the prior.

617 where the sum is taken over all values of x . In communication theory, where the non-
 618 physical uses of entropy began (Gleick 2011), the logarithm used is usually \log_2 but
 619 any logarithm will work and the natural logarithm makes the most sense in population
 620 biology. The classic interpretation of entropy is that it summarizes the uncertainty about
 621 the random variable; it might remind you of diversity indices that we use in ecology (
 622 because, in large part they come from communication theory!).

Now imagine a second discrete random variable Y and assume that the joint probability distribution of X and Y is $f(x, y)$ so that the marginal distributions are $f_x(x) = \sum_y f(x, y)$ and $f_y(y) = \sum_x f(x, y)$. The mutual information of these two random variables is (Cover and Thomas 2006, pg 19ff)

$$I(X; Y) = \sum_{x,y} f(x, y) \log \left[\frac{f(x, y)}{f_x(x) f_y(y)} \right] \quad (86)$$

If $f(x|y)$ is the conditional probability distribution of X given that $Y = y$ then direct computation shows that

$$I(X; Y) = H(X) - H(X|Y) \quad (87)$$

623 where $H(X|Y)$ is computed using Eqn 85 with the conditional distribution $f(x|y)$. The
 624 mutual information is the reduction in the uncertainty of X due to knowledge of Y .

Another quantity associated with entropy is the Kullback-Liebler (KL) distance (Cover and Thomas 2006, pg 19 ff) between two probability distributions $f(x)$ and $g(x)$

$$D(f||g) = \sum_x f(x) \log \left[\frac{f(x)}{g(x)} \right] \quad (88)$$

625 Although $D(f||f) = 0$ this is not a true distance since $D(f||g) \neq D(g||f)$ in general.

626 The advantage of entropy is this: Probability distributions can be hard to visualize and
 627 compare but entropy, information, and KL distance provide scalar summaries of vector
 628 probability distributions.

629 **The Entropy of an Age-Structured Population**

To illustrate use entropy, we consider a simple model for an age-structured population (Mangel 2006) We let $N(a, t)$ denote the number of individuals of age a at time t . If the

rate of natural mortality M is constant, then for all individuals except the birth class

$$N(a, t) = N(a - 1, t - 1)e^{-M} \quad (89)$$

If the fecundity of individuals of age a is ϕ_a the total fecundity of the population at time t is

$$\Phi(t) = \sum_{a>0}^{a_{max}} \phi_a N(a, t) \quad (90)$$

from which the size of the birth class is determined by a density-dependent recruitment function

$$N(0, t) = R(\Phi(t - 1)) \quad (91)$$

When the density dependence of recruitment is such that $R(\Phi)$ is bounded as Φ increases, this population will reach a bounded steady state. If $\bar{N}(a)$ denotes the number of individuals of age a in the steady state and $\bar{\Phi}$ the steady state production of offspring, then

$$\begin{aligned} \bar{N}(0) &= R(\bar{\Phi}) \\ \bar{N}(a) &= \bar{N}(0)e^{-Ma} \end{aligned} \quad (92)$$

From these we compute the fraction of the population of age a and the fraction of reproduction due to individuals of age a

$$p_N(a) = \frac{\bar{N}(a)}{\sum_{a'=0}^{a_{max}} \bar{N}(a')} \quad (93)$$

$$p_\phi(a) = \frac{\bar{N}(a)\phi_a}{\sum_{a'=0}^{a_{max}} \bar{N}(a')\phi_{a'}} \quad (94)$$

To compute the entropy of the distribution $p_N(a)$ we need to do a bit of mathematics, some of which I relegate to the Appendix. For simplicity, let's let A_{max} be so large that we can consider it infinite (in the Appendix you will see how to get around this assumption).

Then we rewrite Eqn 93 as

$$\begin{aligned}
 p_N(a) &= \frac{\bar{N}(0)e^{-Ma}}{\bar{N}(0) + \bar{N}(0)e^{-M} + \bar{N}(0)e^{-2M} + \dots} \\
 &= \frac{e^{-Ma}}{1 + e^{-M} + e^{-2M} + \dots}
 \end{aligned} \tag{95}$$

The first bit of mathematics relegated to the Appendix is that $1 + e^{-M} + e^{-2M} + \dots = \frac{1}{1 - e^{-M}}$. Accepting this, Eqn 95 simplifies to

$$p_N(a) = e^{-Ma}(1 - e^{-M}) \tag{96}$$

This is the geometric distribution. In Appendix 1, you will see that the entropy distribution of population numbers is

$$H_N(M) = \frac{Me^{-M}}{(1 - e^{-M})} - \log(1 - e^{-M}) \tag{97}$$

630 Thus, entropy allows us to summarize in a single number, the spread of the stable age
 631 distribution. In Figure 14, I show the entropy of the stable age distribution as a function
 632 of the rate of natural mortality.

Without further information, we cannot compute the entropy of fecundity. If fecundity at age a is related to mass the allometry $\phi_a = \alpha W(a)^\beta$, then the probability distribution for fecundity becomes

$$p_\phi(a) = \frac{e^{-Ma}W(a)^\beta}{\sum_{a'=0}^{a_{max}} e^{-Ma'}W(a')^\beta} \tag{98}$$

We can also write $W(a) = W_\infty w(a)$ where W_∞ is the asymptotic mass and $w(a)$ captures the time dependent growth of mass, rising from close to zero when $a = 0$ towards 1 as a increases so that

$$p_\phi(a) = \frac{e^{-Ma}w(a)^\beta}{\sum_{a'=0}^{a_{max}} e^{-Ma'}w(a')^\beta} \tag{99}$$

Here we another another advantage of using entropy: we do not need to know asymptotic size (which, since it is asymptotic, we can never know), but rather something about the

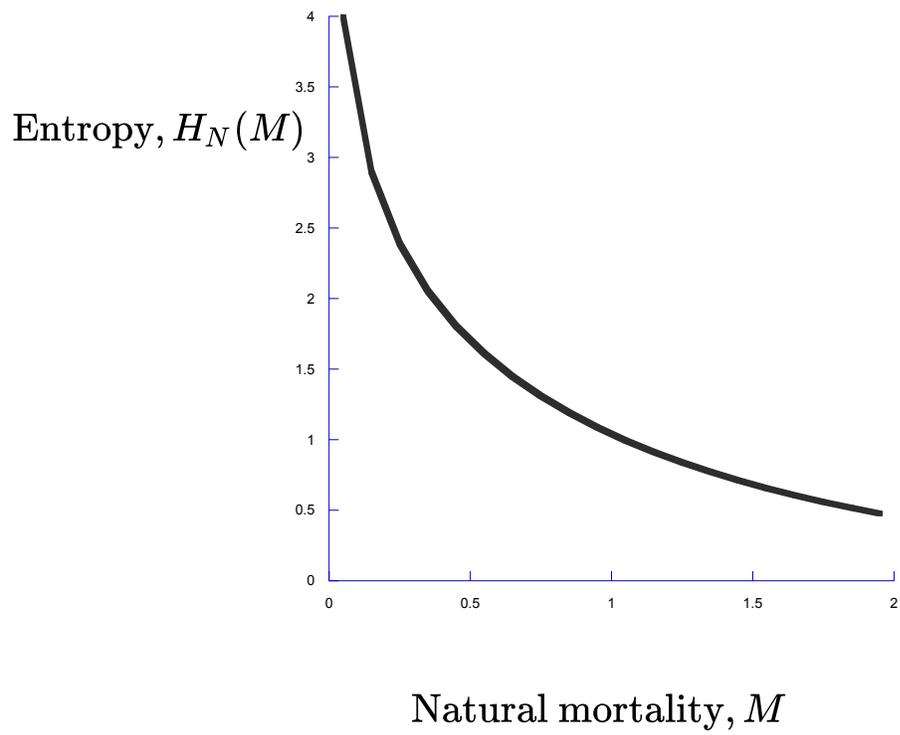


Figure 14: Entropy of an age-structured population as a function of the rate of natural mortality M , showing how a single quantity can capture properties of the entire age structure.

time course of growth. The entropy of fecundity is then

$$H_\phi = - \sum_{a=0}^{a_{max}} p_\phi(a) \log(p_\phi(a)) \quad (100)$$

633 which generally must be computed numerically, unlike that for numbers.

634 **Anthropogenic Mortality and the Effective Rate of Mortality**

To include anthropogenic mortality, we introduce the maximum rate of fishing mortality F and the vulnerability $\nu(a)$ of individuals of age a to fishing, which we model as a logistic curve

$$\nu(a) = \frac{\exp\left[\frac{a-a_{50}}{\sigma_\nu}\right]}{1 + \exp\left[\frac{a-a_{50}}{\sigma_\nu}\right]} \quad (101)$$

635 where a_{50} is the age at which an individual experiences 50% of the maximum fishing
 636 mortality and σ_ν is a measure of how rapidly vulnerability scales with age. It is most
 637 likely that vulnerability is a function of size, rather than age, so we might write $\nu(L(a)) =$

$$\frac{\exp\left[\frac{L(a)-l_{50}}{\sigma_\nu}\right]}{1 + \exp\left[\frac{L(a)-l_{50}}{\sigma_\nu}\right]}$$

With these assumptions, Eqn 92 is replaced by

$$\begin{aligned} \bar{N}(0) &= R(\bar{\Phi}) \\ \bar{N}_F(a) &= \bar{N}(0) e^{-Ma - \sum_{a'=0}^{a-1} \nu(a')F} \end{aligned} \quad (102)$$

The fraction of the population that is age a when fishing mortality is F is

$$p_N(a|F) = \frac{\bar{N}_F(a)}{\sum_{a'>0}^{a_{max}} \bar{N}_F(a')} \quad (103)$$

$$(104)$$

and entropy is

$$H_N(M|F, a_{50}, \sigma_\nu) = - \sum_{a=0}^{a_{max}} p_N(a|F) \log(p_N(a|F)) \quad (105)$$

Since the stable age distributions will differ, we can in principle use the Kullback-Leibler distance to provide a measure of how different they are

$$D(M, F, a_{50}, \sigma_\nu) = \sum_{a=0}^{a_{max}} p_N(a) \log \left[\frac{p_N(a)}{p_N(a|F)} \right] \quad (106)$$

Furthermore, these ideas allow us to introduce the concept of the effective rate of mortality, M_e , which is the rate of natural mortality of a hypothetical population that gives the same entropy as the focal population when it experiences anthropogenic mortality as described above. That is the effective rate of mortality is the solution of the equation

$$H_N(M_e) = H_N(M|F, a_{50}, \sigma_\nu) \quad (107)$$

639 Fishing reduces the age diversity of a population, and this is summarized by the entropy
640 on the right hand side of Eqn 107. By solving for M_e on the left hand side, we are able to
641 say something like “the effect of fishing on this population is equivalent of increasing its
642 rate of natural mortality from M to M_e ”. I encourage you to do some computations with
643 the standard age-structured model and Eqns 106 and 107 to see what kind of intuition
644 about fishing policies based on entropy may emerge.

645 Entropy and Behavior in Nature

646 We close this chapter with a link between individual behavior and population dynamics
647 (Thompson 1986, 1987), concerning the plant *Lomatium dissectum* (Umbelliferae), the
648 moth *Greya subalbata* Braun that lays its eggs in the flower clusters of the plant, and a
649 parasitoid wasp of the *Agathis* genus that attacked the eggs of the moth. When the moth
650 encountered an a cluster, she invariably laid eggs but did not lay in every flower in the
651 cluster.

652 Thompson speculated that the moths behaved in this manner to maximize the uncer-
653 tainty that the parasitoids experienced when they searched for hosts for their own eggs
654 (Thompson 1987, pg 315): “There was no indication that females distinguished among
655 schizocarps within umbellets. Females did not explore several schizocarps [seed pods]

656 before ovipositing, but instead generally began ovipositing in the first schizocarp encoun-
657 tered upon reaching an umbellet [floral cluster]. After ovipositing, a female would either
658 leave the umbellet or move immediately to another schizocarp. The most obvious dif-
659 ference among schizocarps within an umbellet was in the mass of the seeds, which can
660 differ by up to 16-fold within a population (Thompson, 1984). There was no indication,
661 however, that females chose among seed masses within umbellets.” Let us think about
662 this system from the perspective of the parasitoid.

To begin, a simple case is to consider that a particular flower either has a moth egg in
it, with probability $p = 1$, or does not, with probability $p = 0$. In such a case, the entropy
is

$$H(p) = -p \log(p) - (1 - p) \log(1 - p) \quad (108)$$

663 As a way of keeping your calculus skills sharp, you can show that entropy is maximized
664 when $p = 1/2$. That is, the uncertainty about the system is maximized when $p = 1/2$.

665 The actual situation is more complicated, because more than one moth can visit an
666 inflorescence. John and Michael Moody – a mathematical biologist at Washington State
667 University – developed a model for the distribution of eggs across inflorescences and showed
668 that it would be a geometric distribution, truncated to remove 0 (since all schizocarps
669 received an egg) and limited to a maximum number of eggs. In Figure 15, I show a
670 comparison between the observed and predicted distribution of seeds receiving eggs. We
671 have already encountered the entropy of a geometric distribution when we considered
672 the age-structured population model, so you once again see the power of mathematical
673 modeling once again – it connects population dynamics and individual behavior with
674 essentially the same mathematical result. Now, is that sweetness itself or what?

675 **A Look to the Future: What’s Left (Plenty)**

676 Even though this tutorial is long, in a very real sense we have not delved really deeply
677 into the application of Bayesian methods in behavior, population dynamics, and SDLHT.
678 A few other entry points to different literature are these:

- 679 • We have not focussed on detailed physiological mechanism; Ramírez and Marshall
680 (2017) investigate how well natural selection can encode Bayesian priors.
- 681 • We assumed that detection (or predators, prey etc) is perfect but it need not be, so

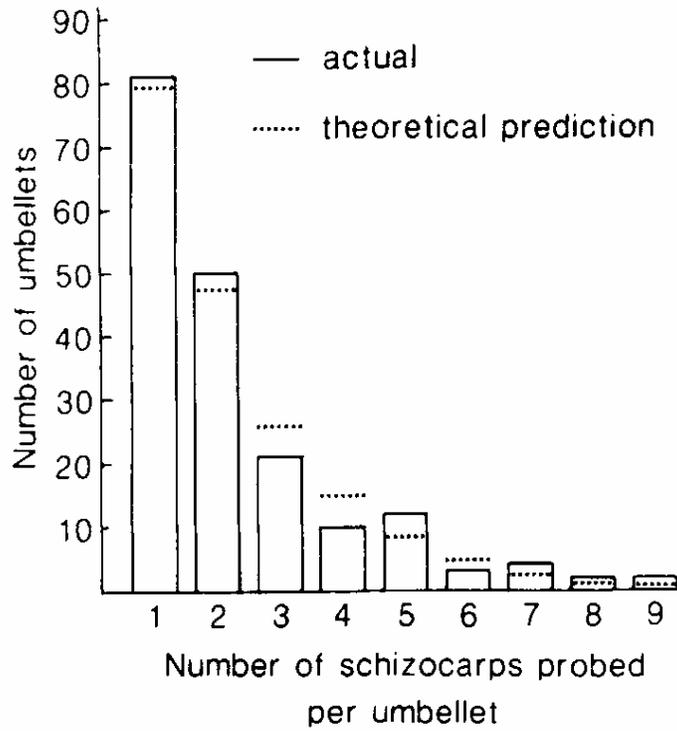


Figure 15: The frequency distribution of the actual number of schizocarps receiving eggs per umbellet from individual female as compared with the distribution predicted by a truncated geometric distribution of movement. From Thompson (1987).

682 that the accuracy of signals needs to be considered (Rosenheim and Mangel 1994,
683 Abbott and Sherratt 2013, McLinn and Stephens 2006, Sherratt 2016, Skelhorn et al
684 2016, Trimmer et al 2017).

- 685 • Aubier et al (2017) use a combination of Bayesian updating and a version of the
686 canonical equation for activity choice to understand the conditions under which
687 un-defended or poorly-defended prey evolve to resemble better-defended prey and
688 whether this mimicry is parasitic (quasi-Batesian) or mutualistic (Müllerian). They
689 compare a learning and optimally behaving predator with a Pavlovian model based
690 on conditioning and conclude that if predators forage efficiently to maximize their
691 long-term payoff, genuine quasi-Batesian mimicry will be rare, which is the case in
692 nature.
- 693 • Although we have treated information as a property of individuals, there are cases
694 where Bayesian analysis is appropriate for the use of social information (Barbier and
695 Watson 2016, Gil et al 2017, 2018, Lee et al 2016).
- 696 • We only touched on the role of Bayesian analysis in mate choice; there are rich op-
697 portunities for application (Luttbeg 1996, 2002, Luttbeg 2004, Luttbeg and Langen
698 2004, Rands et al 2011, Castellano 2015).
- 699 • Bayesian methods are at the core of adaptive management – in which we use man-
700 agement action to both obtain something from a system and learn about that system
701 (Kuikka et al 2014, Rosenheim et al 2011, Williams 2015, Meisner et al 2016, Chades
702 et al 2017).
- 703 • We also only touched on the ways that entropy can be used. Donaldson-Matasci et
704 al (2010, 2013) discuss ways integrate entropy and optimization into a life history
705 context.
- 706 • Especially in constant environments, the posterior will converge to a nearly steady
707 state distribution that may vary depending on experience. This allows the devel-
708 opment of behavioral syndromes and animal personality (Sih et al 2004, Biro and
709 Stamps 2008, Luttbeg and Sih 2010, Patrick and Weimerskirch 2014).

710 **Appendix 1: Details About the Entropy of Population Numbers**

In this Appendix, we will fill in the details about entropy of the distribution of population numbers, $H_N(M)$ given in Eqn 97. To begin, we need to show that $1 + e^{-M} + e^{-2M} + \dots = \frac{1}{1-e^{-M}}$. To do this, let us denote by $\mathcal{S}(a-1)$ the sum of the first $a-1$ terms

$$\mathcal{S}(a-1) = 1 + e^{-M} + e^{-2M} + \dots + e^{-M(a-1)} \quad (109)$$

Now multiply both sides by e^{-M}

$$e^{-M}\mathcal{S}(a-1) = e^{-M} + e^{-2M} + e^{-3M} \dots + e^{-Ma} \quad (110)$$

Subtracting Eqn 110 from Eqn 109, and recognizing that most terms cancel, we obtain

$$\mathcal{S}(a-1) - e^{-M}\mathcal{S}(a-1) = 1 - e^{-Ma} \quad (111)$$

so that

$$\mathcal{S}(a-1) = \frac{1 - e^{-Ma}}{1 - e^{-M}} \quad (112)$$

If we let a become continually larger, $\mathcal{S}(a-1)$ approaches $1 + e^{-M} + e^{-2M} + \dots$ and the right hand side of Eqn 112 approaches $\frac{1}{1-e^{-M}}$.

We are now ready to tackle the entropy itself. Since the fraction of the population that is age a is $p_N(a) = e^{-Ma}(1 - e^{-M})$ entropy is

$$H_N(a) = - \sum_a e^{-Ma}(1 - e^{-M}) \log \left[e^{-Ma}(1 - e^{-M}) \right] \quad (113)$$

Expanding the logarithm allows us to write

$$H_N(a) = - \sum_a e^{-Ma}(1 - e^{-M}) \log(e^{-Ma}) - \sum_a e^{-Ma}(1 - e^{-M}) \log(1 - e^{-M}) \quad (114)$$

The second sum is actually easier to deal with than the first, so let's dispose of it promptly. Ignoring the negative sign, we first take the terms that do not depend upon age out of

the summation to obtain

$$\sum_a e^{-Ma}(1 - e^{-M})\log(1 - e^{-M}) = (1 - e^{-M})\log(1 - e^{-M}) \sum_a e^{-Ma} \quad (115)$$

and since $\sum_a e^{-Ma} = \frac{1}{1 - e^{-M}}$, the second sum is simply $\log(1 - e^{-M})$.

To deal with the first sum in Eqn 114, let's again forget about the negative sign. The first thing we do is take the logarithm (which brings down a negative sign) and the second thing we do is take things that do not depend upon age out of the summation, so that

$$\begin{aligned} \sum_a e^{-Ma}(1 - e^{-M})\log(e^{-Ma}) &= \sum_a e^{-Ma}\log(1 - e^{-M})(-Ma) \\ &= -M(1 - e^{-M}) \sum_a ae^{-Ma} \end{aligned} \quad (116)$$

Now remembering that the derivative $\frac{d}{dM}(e^{-Ma}) = -ae^{-Ma}$, we follow this sequence of steps

$$\sum_a ae^{-Ma} = - \sum_a \frac{d}{dM}(e^{-Ma}) = - \frac{d}{dM} \sum_a e^{-Ma} = - \frac{d}{dM}(1 - e^{-M})^{-1} \quad (117)$$

We are almost there! We take the derivative on the right hand side of Eqn 117 remembering the chain rule

$$- \frac{d}{dM}(1 - e^{-M})^{-1} = \frac{e^{-M}}{(1 - e^{-M})^2} \quad (118)$$

so that we conclude that $\sum_a ae^{-Ma} = \frac{e^{-M}}{(1 - e^{-M})^2}$, which we use in the right hand side of Eqn 116 and then back in Eqn 114 to obtain

$$H_N(M) = \frac{Me^{-M}}{1 - e^{-M}} - \log(1 - e^{-M}) \quad (119)$$

711 as the entropy of the distribution of numbers in the age-structured population.

712 **Appendix 2: The Normal-Normal Model and Updating Informa-**
713 **tion About Temperature**

714 In this Appendix, we explore using a normal prior distribution for an unknown mean of
715 a normal distribution, keeping updating about temperature in mind as an example. The
716 development of the model is not difficult but does require careful attention to algebra.

717 We imagine an individual in ambient temperature T_a , which is not observed perfectly.
718 For each of K days, the individual observes the temperature with T_k denoting the observed
719 temperature on day k . Thus the vector of observations is $\vec{v}(T_K) = (T_1, T_2, \dots, T_K)$.

Next let us assume that these daily observations of temperature are normally distributed with the ambient temperature as the mean and an observational standard deviation σ_{obs} . Thus, the probability density associated with a single observation is $\frac{1}{\sqrt{2\pi}\sigma_{obs}} \exp\left[-\frac{1}{2\sigma_{obs}^2}(T_k - T_a)^2\right]$ and if the observations are independent, then the probability of the K days of observations is

$$Pr\{\vec{v}(T_K)|T_a, \sigma_{obs}\} = \left[\frac{1}{\sqrt{2\pi}\sigma_{obs}}\right]^K \exp\left[-\frac{1}{2\sigma_{obs}^2}\sum_{k=1}^K(T_k - T_a)^2\right] \quad (120)$$

Our next step is to manipulate the summation on the right hand side; I am going to lay out all of the steps,. To do this, we let $\bar{T}_K = \frac{1}{K}\sum_{k=1}^K T_k$ denote the mean of the temperature observations and write

$$\begin{aligned} \sum_{k=1}^K(T_k - T_a)^2 &= \sum_{k=1}^K(T_k - \bar{T}_K - T_a + \bar{T}_K)^2 \\ &= \sum_{k=1}^K[(T_k - \bar{T}_K) - (T_a - \bar{T}_K)]^2 \\ &= \sum_{k=1}^K(T_k - \bar{T}_K)^2 - 2 \cdot \sum_{k=1}^K(T_k - \bar{T}_K)(T_a - \bar{T}_K) \\ &\quad + \sum_{k=1}^K(T_a - \bar{T}_K)^2 \end{aligned}$$

The sample variance is $s_K^2 = \frac{1}{K}\sum_{k=1}^K(T_k - \bar{T}_K)^2$ and we recognize that with the summations involving $(T_a - \bar{T}_K)$ that expression can be factored out of the summation since it does not depend upon k . We then obtain

$$\sum_{k=1}^K(T_k - T_a)^2 = Ks_K^2 + K(\bar{T}_K - T_a)^2 - 2(T_a - \bar{T}_K)\sum_{k=1}^K(T_k - \bar{T}_K)$$

Now $\sum_{k=1}^K(T_k - \bar{T}_K) = \sum_{k=1}^K T_k - \sum_{k=1}^K \bar{T}_K = K\bar{T}_K - K\bar{T}_K = 0$. The point of all this

analysis is that we can rewrite Eqn 120 as

$$Pr\{\vec{v}(T_K)|T_a, \sigma_{obs}\} = \left[\frac{1}{\sqrt{2\pi}\sigma_{obs}} \right]^K \exp \left[- \frac{1}{2\sigma_{obs}^2} \{Ks_K^2 + K(\bar{T}_K - T_a)^2\} \right] \quad (121)$$

For Bayesian analysis, we want to interpret Eqn 121 as the likelihood $\mathcal{L}(T_a|\bar{T}_K, \sigma_{obs})$ of T_a given the data and the observation uncertainty (Hilborn and Mangel 1997, Mangel 2006), which means that anything not depending on T_a can be treated as a constant. We absorb these into an anonymous proportionality constant and thus write

$$\mathcal{L}(T_a|\bar{T}_K, \sigma_{obs}) \propto \exp \left[- \frac{1}{2\sigma_{obs}^2} K(\bar{T}_K - T_a)^2 \right] \quad (122)$$

Eqn 122 will be the likelihood for the Bayesian analysis. Next we need a prior distribution for T_a .

To begin, let us assume that the long-term, evolutionarily experienced average temperature is \bar{T}_e and that in any particular year the ambient temperature is drawn from a normal distribution centered around this average with environmental variance σ_e^2 . Thus the prior for T_a is

$$f_{prior}(T_a|T_e) = \frac{1}{\sqrt{2\pi}\sigma_e} \exp \left[- \frac{1}{2\sigma_e^2} (T_a - T_e)^2 \right] \quad (123)$$

Consequently the posterior distribution for ambient temperature given the data is

$$f_{post}(T_a|\bar{T}_K, \sigma_{obs}, T_e) \propto \mathcal{L}(T_a|\bar{T}_K, \sigma_{obs}) f_{prior}(T_a|T_e) \quad (124)$$

We multiply the likelihood and prior, taking advantage of the rules for multiplying exponentials, and obtain the posterior

$$f_{post}(T_a|\bar{T}_K, \sigma_{obs}, T_e) \propto \exp \left[- \frac{1}{2\sigma_{obs}^2} K(\bar{T}_K - T_a)^2 - \frac{1}{2\sigma_e^2} (T_a - T_e)^2 \right] \quad (125)$$

We now manipulate the argument of the exponential in Eqn 125

$$\begin{aligned} \frac{1}{2\sigma_{obs}^2} K(\bar{T}_K - T_a)^2 + \frac{1}{2\sigma_e^2} (T_a - T_e)^2 &= \frac{1}{2\sigma_{obs}^2} K[\bar{T}_K^2 - 2T_a\bar{T}_K + T_a^2] \\ &\quad + \frac{1}{2\sigma_e^2} [T_a^2 - 2T_aT_e + T_e^2] \end{aligned}$$

Keeping our eyes on the prize, let's now collect terms according to power of T_a , since that is the focus of our analysis:

$$\begin{aligned} \frac{1}{2\sigma_{obs}^2}K(\bar{T}_K - T_a)^2 + \frac{1}{2\sigma_e^2}(T_a - T_e)^2 &= \frac{T_a^2}{2} \left[\frac{K}{\sigma_{obs}^2} + \frac{1}{\sigma_e^2} \right] - T_a \left[\frac{K\bar{T}_K}{\sigma_{obs}^2} + \frac{T_e}{\sigma_e^2} \right] \\ &\quad + \frac{\bar{T}_K^2}{2\sigma_{obs}^2} + \frac{T_e^2}{2\sigma_e^2} \end{aligned}$$

Now the third and fourth terms on the right-hand side above do not depend upon T_a which means that we can absorb them into the normalization constant in the posterior density for T_a . Hence we can rewrite Eqn 125 as

$$f_{post}(T_a|\bar{T}_K, \sigma_{obs}, T_e) \propto \exp \left[-\frac{1}{2} \left\{ T_a^2 \left(\frac{K}{\sigma_{obs}^2} + \frac{1}{\sigma_e^2} \right) - 2T_a \left(\frac{K\bar{T}_K}{\sigma_{obs}^2} + \frac{T_e}{\sigma_e^2} \right) \right\} \right] \quad (126)$$

Suppose we write the posterior as as a Gaussian with mean μ_K and standard deviation σ_K , where K reminds us that K observations have been collected. Then it would be

$$\begin{aligned} f_{post}(T_a|\bar{T}_K, \sigma_{obs}, T_e) \propto \exp \left[-\frac{1}{2\sigma_K^2}(T_a - \mu_K)^2 \right] = \\ \exp \left[-\frac{1}{2} \left(\frac{T_a^2}{\sigma_K^2} - 2\frac{\mu_K T_a}{\sigma_K^2} + \mu_K^2 \right) \right] \end{aligned} \quad (127)$$

We now compare the right hand sides of Eqns 126 and 127. The terms involving T_a^2 will match if we set

$$\frac{1}{\sigma_K^2} = \frac{K}{\sigma_{obs}^2} + \frac{1}{\sigma_e^2} \quad (128)$$

which means that

$$\sigma_K^2 = \frac{1}{\frac{K}{\sigma_{obs}^2} + \frac{1}{\sigma_e^2}} = \frac{\sigma_{obs}^2 \sigma_e^2}{K\sigma_e^2 + \sigma_{obs}^2} \quad (129)$$

That is, the variance in our posterior distribution after K observations is obtained by weighting the observation and environmental variances according to the number of observations.

A formula such as Eqn 128 or 129 has lead to the very helpful and intuitive introduction of the notion of precision, defined to be the reciprocal of the variance. We will use τ

for precision with subscripts denoting posterior, observation, or environmental variables. In that case Eqn 128 becomes

$$\tau_K = K\tau_{obs} + \tau_e \quad (130)$$

This is very nice. It tells us that the precision with which we know the posterior is the sum of the observational precision, taking into account the number of observations, and the environmental precision. Comparing the terms in Eqns 126 and 127 that involve T_a we conclude

$$\frac{\mu_K}{\sigma_K^2} = \frac{K\bar{T}}{\sigma_{obs}^2} + \frac{T_e}{\sigma_e^2} \quad (131)$$

and if we use our new friend precision, Eqn 131 can be rewritten as

$$\mu_K = \frac{1}{\tau_K}(\tau_{obs}K\bar{T}_K + \tau_e T_e) \quad (132)$$

I'd like to do one more un-numbered series of manipulations that will allow us to interpret μ_k . Since $\tau_K = K\tau_{obs} + \tau_e$ we can write

$$\begin{aligned} \mu_K &= \frac{\tau_{obs}K\bar{T}_K + \tau_e T_e}{K\tau_{obs} + \tau_e} \\ &= \bar{T}_K \left(\frac{K\tau_{obs}}{K\tau_{obs} + \tau_e} \right) + T_e \left(\frac{\tau_e}{K\tau_{obs} + \tau_e} \right) \end{aligned}$$

so that μ_K is the weighted mixture of the observations (\bar{T}_K) and evolutionary information (T_e) with the weighting determined by the precisions (reciprocal variances) and the amount of data collected. This is again sweetness itself. In summary, we now write the posterior as

$$\begin{aligned} f_{post}(T_a|\bar{T}_K, \tau, T_e, \tau_e) &\propto \sqrt{\frac{\tau_K}{2\pi}} \exp \left[-2\tau_K(T_a - \mu_K)^2 \right] \\ \tau_K &= K\tau_{obs} + \tau_e \\ \mu_K &= \frac{1}{\tau_K}(\tau_{obs}K\bar{T}_K + \tau_e T_e) \end{aligned} \quad (133)$$

Note that if $K = 1$ and we let T_{obs} denote the observed temperature, then we have

$$\begin{aligned}\tau_1 &= \tau_{obs} + \tau_e \\ \mu_1 &= \frac{\tau_{obs}T_{obs} + \tau_e T_e}{\tau_{obs} + \tau_e}\end{aligned}\tag{134}$$

720 and these give us the updating formula. At any time, the organism has a prior distribution
721 for T_a with mean μ_0 and precision τ_0 and collects one observation T_{obs} ; the posterior mean
722 and precision are then $\mu_1 = \frac{\tau_{obs}T_{obs} + \tau_0\mu_0}{\tau_{obs} + \tau_0}$ and $\tau_1 = \tau_{obs} + \tau_0$ which become the prior
723 parameters before the next observation is collected.

We need to do one more calculation before we could derive a SDP model. That is, given current estimates for the environment – summarized by μ_0 and $\tau_0 = 1/\sigma_0$, what is the distribution of possible observed temperatures? This is called the posterior predictive density, since we are predicting a temperature given the posterior parameters. Let us denote this density by $f_{pre}(T_{obs}|\mu_K, \sigma_K)$; it will be a normal density for T_{obs} conditioned on T_a times a normal density for T_a given the posterior parameters (rather than the prior parameters) summed (i.e. integrated) over all possible values of T_a , so that we have

$$f_{pre}(T_{obs}|\mu_K, \sigma_K) \propto \int \exp\left[-\frac{(T_{obs} - T_a)^2}{2\sigma_{obs}^2}\right] \exp\left[-\frac{(T_a - \mu_K)^2}{2\sigma_K^2}\right] dT_a \tag{135}$$

724 Now, we could once again do the work of completing the square as above, and as you
725 might surmise, the result will be that the posterior predictive distribution will be a normal
726 distribution. I am not going to do that, although I suggest that you try it yourself or
727 to implement it as we did with posteriors in a changing world. Rather, we are going to
728 use the following argument. Note that $T_{obs} = (T_{obs} - T_a) + T_a$ and that $(T_{obs} - T_a)$ is
729 a normal distribution with mean 0 and standard deviation σ_{obs} and that T_a is a normal
730 distribution with mean μ_k and standard deviation σ_k . For two independent random
731 variables, Z_1 and Z_2 the mean of the sum is $\mathcal{E}(Z_1 + Z_2) = \mathcal{E}(Z_1) + \mathcal{E}(Z_2)$ and the variance
732 is $Var(Z_1 + Z_2) = Var(Z_1) + Var(Z_2)$. Hence we conclude without any algebra that
733 $f_{pre}(T_{obs}|\mu_K, \sigma_K)$ is a normal distribution with mean μ_K and variance $\sigma_{obs}^2 + \sigma_K^2$.

734 Literature Cited

735 Abbott, K.R. and T.N. Sherratt. 2013. Optimal sampling and signal detection: unifying
736 models of attention and speed-accuracy trade-offs. Behavioral Ecology doi:10.1093/beheco/art001

737 Abramowitz, M. and I. Stegun. 1965. *Handbook of Mathematical Functions* (Wash-
738 ington, Superintendent of Documents). Republished by Dover Publications

739 Alonzo, S.H. and R.R. Warner. 2000. Dynamic games and field experiments exam-
740 ining intra- and intersexual conflict: explaining counterintuitive mating behavior in a
741 Mediterranean wrasse, *Symphodus ocellatus*. Behavioral Ecology 11: 56-70

742 Alonzo, S.H. and K.L. Heckman. 2010. The unexpected but understandable dynamics
743 of mating, paternity and paternal care in the ocellated wrasse. Proceedings of the Royal
744 Society of London B 277: 115-122

745 Arnoldini, M., Mostowy, R., Bonhoeffer, S., and M. Ackermann M. 2012. Evolution of
746 Stress Response in the Face of Unreliable Environmental Signals. PLoS Computational
747 Biology 8(8): e1002627. doi:10.1371/journal.pcbi.1002627

748 Aubier, T.G., Joron, M., and T.N. Sherratt. 2017. Mimicry among unequally defended
749 prey should be mutualistic when predators sample optimally. The American Naturalist
750 189: 267-282

751 Barbier, M and J.R. Watson. 2016. The spatial dynamics of predators and the ben-
752 efits and costs of sharing information. PLoS Computational Biology 12(10): e1005147.
753 doi:10.1371/journal.pcbi.1005147

754 Barrett, L. F. 2020. *71/2 Lessons About the Brain*. Houghton Mifflin Harcourt, Boston
755 and New York.

756 Bayes, T. 1763. An Essay towards solving a problem in the doctrine of chances.
757 Philosophical Transactions (1683-1775) 53: 370-418

758 Biro, P.A. and J.A. Stamps. 2008. Are animal personality traits linked to life history
759 productivity? Trends in Ecology and Evolution 23: 361-368

760 Boettiger C., Mangel M., and S. Munch. 2015. Avoiding tipping points in fisheries
761 management through Gaussian process dynamic programming. Proceedings of the Royal
762 Society B 282: 20141631. <http://dx.doi.org/10.1098/rspb.2014.1631>

763 Castellano, S. 2015. Bayes's rule and bias roles in the evolution of decision making.
764 Behavioral Ecology 26: 282-292

765 Cover, T.M. and Thomas, J.A. 2006. Elements of Information Theory. 2nd edn
766 (Wiley-Interscience) Hoboken, NJ

767 Cane, V.R. 1959. Behaviour Sequences as Semi-Markov Chains. Journal of the Royal
768 Statistical Society. Series B (Methodological) 21: 36-58

769 Chades, I., Nicols, S., Rout, T.M., Péron, M., Dujardin, Y., Pichancourt, J-B., Hast-
770 ings, A., and C.E. Hauser. 2017. Optimization methods to solve adaptive management

771 problems. *Theoretical Ecology* 10 DOI 10.1007/s12080-016-0313-0

772 Clark, C.W. and M. Mangel. 2000. *Dynamic State Variable Models in Ecology. Methods and Applications*
773 Oxford University Press, Oxford, UK.

774 de Jong, M.A., Kesbeke, F. M. N. H. , Brakefield, P. M. , and B. J. Zwaan. 2010.
775 Geographic variation in thermal plasticity of life history and wing pattern in *Bicyclus*
776 *anyana* *Climate Research* 43:91-102

777 Dominy, N.J., Yeakel, J.D., Bhat,U., Ramsden,L., Wrangham, R.W., and P.W. Lucas.
778 2016. How chimpanzees integrate sensory information to select figs. *Interface Focus* 6:
779 20160001. <http://dx.doi.org/10.1098/rsfs.2016.0001>

780 Donaldson-Matasci, M.C., Bergstrom, C.T., and M. Lachman. 2010. The fitness value
781 of information. *Okios* : 219-230

782 Donaldson-Matasci, M.C., Bergstrom, C.T., and M. Lachman. 2013. When unreliable
783 cues are good enough. *The American Naturalist* 182: 313-327

784 Dorazio, R.M. 2016. Bayesian data analysis in population ecology: motivations, meth-
785 ods, and benefits. *Population Ecology* 58: 31-44

786 Efron, B. 1986. Why Isn't Everyone a Bayesian? *The American Statistician* 40:1-5

787 Eliassen, S., Jørgensen, C., Mangel, M. and J. Giske. 2007. Exploration or exploita-
788 tion: life expectancy changes the value of learning in foraging strategies. *Oikos* 116:513-
789 523

790 Eliassen, S., Jørgensen, C., Mangel, M. and J. Giske. Quantifying the adaptive value
791 of learning in foraging behavior. *American Naturalist* (2009) doi:10.1086/605370

792 English, S., Fawcett, T.W., Higgison, A.D., Trimmer, P.C., and T. Uller. 2016. Adap-
793 tive use of information during growth can explain long-term effects of early life experiences.
794 *American Naturalist* 187: DOI: 10.1086/685644

795 Frankenhuis, W.E. and K. Panchanathan. 2011. Balancing sampling and specializa-
796 tion: an adaptationist model of incremental development *Proceedings of the Royal Society*
797 *B* 278 doi: 10.1098/rspb.2011.005

798 Frankenhuis, W., Panchanathan, K., and A.G. Barto. 2018. Enriching behavioral ecol-
799 ogy with reinforcement learning methods. *Behavioural Processes*. <https://doi.org/10.1016/j.beproc.2018.01>

800 Gelman, A., Carlin, J.B., Stern, H.S, and D.B. Rubin. 2016. *Bayesian Data Analysis*.
801 (Boca Raton, FL Chapman & Hall/CRC)

802 Gil, M.A., Emberts, Z., Jones, H., and C.M. St. Mary. 2017. Social information on
803 fear and food drives animal grouping and fitness. *The American Naturalist* 189: 227-241

804 Gil, M.A., Hein, A.M., Spiegel, O., Baskett, M.L., and A. Sih. 2018. Social information

805 links individual behavior to population and community dynamics. Trends in Ecology and
806 Evolution doi.org/10.1016/j.tree.2018.04.010

807 Gleick, J. 2011. *The Information. A History. A Theory. A Flood.* Pantheon Books,
808 New York

809 Hammersley, J. M. 1971. "No matter, Never mind!" Bulletin of the Institute of Math-
810 ematics and Its Applications 7(12)

811 Hammersley, J. M. 1973. How is Research Done? Bulletin of the Institute of Mathe-
812 matics and Its Application, July 1973: 214-215

813 Hammersley, J. M. 1974. Poking about for the vital juices of mathematical research.
814 Bulletin of the Institute of Mathematics and Its Applications , July/August 1974:235-247

815 Hilborn, R. and M. Mangel. 1997. *The Ecological Detective. Confronting Models with*
816 *Data.* Princeton University Press, Princeton, NJ.

817 Isaacson, W. 2014. *The Innovators: How a Group of Hackers, Geniuses, and Geeks*
818 *Created the Digital Revolution* (Simon and Schuster, New York)

819 Iwasa, Y., Suzuki, Y., and H. Matsuda. 1984. The theory of oviposition strategy
820 of parasitoids. I. Effect of mortality and limited egg number. Theoretical Population
821 Biology 26: 205-227

822 Jaynes, E.T. 2003. *Probability Theory. The Logic of Science.* Cambridge University
823 Press, Cambridge, UK

824 Knight, F.H. *Risk, Uncertainty, and Profit*(University of Chicago Press, Chicago, IL,
825 1921/1971)

826 Krebs, J.R., Kacelnik, A., and P. Taylor. 1978. Test of optimal sampling by foraging
827 great tits. Nature 275: 27-31

828 Kuikka, S., Vanhatalo, J., Pulkkinen, H., Mäntyniemi, S. and J. Corander. 2014.
829 Experiences in Bayesian inference in Baltic salmon management. Statistical Science 29:
830 42-49

831 Lee, A.E.G., Ounsley, J.P., Coulson, T., Rowcliffe, J.M., and G. Cowlishaw G. 2016.
832 Information use and resource competition: an integrative framework. Proceedings of the
833 Royal Society B 283: 20152550. <http://dx.doi.org/10.1098/rspb.2015.2550>

834 Lehtonen, J. 2016. The Lambert W function in ecological and evolutionary models.
835 Methods in Ecology and Evolution 7: 1110-1118

836 Lloyd, K. and D. S. Leslie. 2013. Journal of the Royal Society Interface 10: 20130069.
837 <http://dx.doi.org/10.1098/rsif.2013.0069>

838 Luttbeg, B. 1996. A comparative Bayes tactic for mate assessment and choice. Be-

839 havioral Ecology 7: 451-460

840 Luttbeg, B. 2002. Assessing the robustness and optimality of alternative decision rules
841 with varying assumptions. *Animal Behavior* 63:805-814

842 Luttbeg, B. 2004. Female mate assessment and choice behavior affect the frequency
843 of alternative male mating tactics. *Behavioral Ecology* 15: 239-247 DOI: 10.1093/behe-
844 co/arh002

845 Luttbeg, B. and O.J. Schmitz. 2000. Predator and prey models with flexible individual
846 behavior and imperfect information. *American Naturalist*, 155: 669-683.

847 Luttbeg, B. and T. Langen. 2004. Comparing alternative models to empirical data:
848 Cognitive models of western scrub-jay foraging behavior. *The American Naturalist* 163:
849 263-276

850 Luttbeg, B. and A. Sih. 2010. Risk, resources and state-dependent adaptive be-
851 havioural syndromes. *Philosophical Transactions of the Royal Society B* 365: 3977-3990
852 doi:10.1098/rstb.2010.0207

853 Luttbeg, B. and G.C.Trussell. 2013. How the informational environment shapes how
854 prey estimate predation risk and the resulting indirect effects of predators. *American*
855 *Naturalist* 181:182- 194

856 Luttbeg, B. and R.R. Warner. 1999. Reproductive decision-making by female pea-
857 cock wrasses: flexible versus fixed behavioral rules in variable environments. *Behavioral*
858 *Ecology* 10:666-674

859 Mangel, M. 1990. Dynamic information in uncertain and changing worlds, *Journal of*
860 *Theoretical Biology*, 146:317-332

861 Mangel, M. 1994. Barrier transitions driven by fluctuations, with applications to
862 ecology and evolution. *Theoretical Population Biology*. 45:16-40

863 Mangel, M. 2006. *The Theoretical Biologist's Toolbox. Quantitative Methods for Ecol-*
864 *ogy and Evolutionary Biology*. Cambridge University Press, Cambridge, UK

865 Mangel, M. Climate change and salmonid life history variation. 1994a. *Deep Sea*
866 *Research, II (Topical Studies in Oceanography)* 41:75-106

867 Mangel, M. 2015. Stochastic Dynamic Programming illuminates the link between
868 environment, physiology, and evolution. *Bulletin of Mathematical Biology*. 77:857-877

869 Mangel, M. 2020a. From Demography (R_0) to State Dependent Life History Theory
870 (SDLHT): A Tutorial. Available at my web site (probably where you got this document).

871 Mangel, M. 2020b. Free for the Doing: Some Potential Projects Using State Depen-
872 dent Life History Theory. Available at my web site (also probably where you got this

873 document).

874 Mangel, M. 2006. *The Theoretical Biologist's Toolbox. Quantitative Methods for Ecology and Evolutionary Biology*. Cambridge University Press, Cambridge, UK.

875
876 Mangel, M. and D. Ludwig. 1977. Probability of extinction in a stochastic competition. Society for Industrial and Applied Mathematics (SIAM) Journal on Applied
877 Mathematics 33:256-266.

878
879 Mangel, M. and C.W. Clark. 1983. Uncertainty, search, and information in fisheries. Journal of the International Council for the Exploration of the Seas 43:93-103

880
881 Mangel, M. and J.H. Beder. 1985. Search and stock depletion: theory and applications. Canadian Journal of Fisheries and Aquatic Sciences 42:150-163

882
883 Mangel, M. and B.D. Roitberg. 1989. Dynamic information and host acceptance by a tephritid fruit fly. Ecological Entomology 14:181-189

884
885 Marshall, H.H., Carter, A.J., Ashford, A., Rowcliffe, J.M. and G. Cowlishaw. 2013. How do foragers decide when to leave a patch? A test of alternative models under natural
886 and experimental conditions. Journal of Animal Ecology doi: 10.1111/1365-2656.12089

887
888 Martz, H.F. and R.A. Waller. 1982. *Bayesian Reliability Analysis* (New York, Wiley Interscience)

889
890 Mathot, K.J. and S.R.X. Dall. 2013. Metabolic rates can rive individual differences in information and insurance use under the risk of starvation. The American Naturalist
891 182: 611-620

892
893 McElreath, R. 2016. *Statistical Rethinking*. Chapman and Hall, Boca Raton, FL

894
895 McLinn, C.M. and D.W. Stephens. 2006. What makes information valuable: signal reliability and environmental uncertainty. Animal Behaviour. 71:1119-1129

896
897 McNamara, J.M. and S.R.X. Dall. 2010. Information is a fitness enhancing resource. Oikos 119: 231-236

898
899 Nettle, D., Frankenhuys, W.E., and I.J. Rickard. 2013. The evolution of predictive adaptive responses in human life history. Proceedings of the Royal Society B 280: 20131343. <http://dx.doi.org/10.1098/rspb.2013.1343>

900
901 Nettle D. and M. Bateson. 2015. Adaptive developmental plasticity: what is it, how can we recognize it and when can it evolve? Proceedings of the Royal Society B 282: 20151005. <http://dx.doi.org/10.1098/rspb.2015.1005>

902
903 Panchanathan, K. and W.E. Frankenhuys. 2016. The evolution of sensitive periods in a model of incremental development. Proceedings of the Royal Society B 283: 20152439. <http://dx.doi.org/10.1098/rspb.2015.2439>

907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940

Patrick, S.C. and H. Weimerskirch, H. 2014. Personality, foraging and fitness consequences in a long lived seabird. PLoS ONE 9: e87269.

Pike, R.K., McNamara, J.M., and A.I. Houston. 2016. A general expression for the reproductive value of information. Behavioral Ecology 27:1296-1303. doi:10.1093/beheco/arw044

Powell, W.B. 2011. *Approximate Dynamic Programming. Solving the Curves of Dimensionality* (New York, Wiley Interscience)

Powell, W.B. 2014. Clearing the jungle of stochastic optimization. In INFORMS Tutorials in Operations Research. Published online: 27 Oct 2014; 109-137. <http://dx.doi.org/10.1287/educ.2014>

Powell, W.B. 2015a. Tutorial on stochastic optimization in energy, I: Modeling and policies. IEEE Transactions on Power Systems

Powell, W.B. 2015b. Tutorial on stochastic optimization in energy, II: An energy storage illustration. IEEE Transactions on Power Systems

Powell, W.B., Simao, H.P., and Bouzaiene-Ayari, B. 2012. Approximate dynamic programming in transportation and logistics: a unified framework EURO J Transp Logist 1:237-284

Puterman, M.L. 1994. *Markov Decision Processes. Discrete Stochastic Dynamic Programming*. Wiley Interscience, New York

Railsback, S.F. and B.C. Harvey. 2020. *Modeling Populations of Adaptive Individuals*. Princeton University Press, Princeton, NJ

Ramírez, J.C. and J.A.R. Marshall. 2017. Can natural selection encode Bayesian priors? Journal of Theoretical Biology 426: 57-66

Rands, S.A., Evans, M.R., and R.A. Johnstone R.A. 2011. The Dynamics of honesty: Modelling the growth of costly, sexually-selected ornaments. PLoS ONE 6(11): e27174. doi:10.1371/journal.pone.0027174

Rosenheim, J.A. and M. Mangel. 1994. Patch-leaving rules for parasitoids with imperfect host discrimination. Ecological Entomology 19:374-380

Rosenheim, J.A., Parsa, S., Forbest, A.A., Krimmel, W.A., Law, Y.H., Segoli, M., Segoli, M., Sivakoff, F.S., Zaviezo, T. and K. Gross. 2011. Econinformatics for integrated pest management: Expanding the applied insect ecologist's tool-kit. Journal of Economic Entomology 104:331-342

Satterthwaite, W.H., Beakes, M.P., Collins, E., Swank, D.R., Merz, J.E., Titus, R.G., Sogard, S.M., and M. Mangel. 2009. Steelhead life history on California's Central Coast: Insights from a State-Dependent Model. Transactions of the American Fisheries Society 132:532-548.

941 Schmidt, K.A., Dall, S.R.X., and J.A. van Gils. 2010. The ecology of information: an
942 overview on the ecological significance of making informed decisions. *Oikos* 119: 304-316

943 Sherratt, T.N. 2016. State-dependent optimization: a comment on Sklehorn et al.
944 *Behavioral Ecology* doi:10.1093/beheco/arw02

945 Skelhorn, J., Halpin, C.G. and C. Rowe. 2016. Learning about aposematic prey.
946 *Behavioral Ecology* doi:10.1093/beheco/arw009

947 Stamps, J.A. and V.V. Krishnan. 2014. Combining information from ancestors and
948 personal experiences to predict individual differences in developmental trajectories. *Amer-*
949 *ican Naturalist* 184: 647-657

950 Stamps, J.A. and W. Frankenhuis. 2016. Bayesian models of development. *Trends in*
951 *Ecology and Evolution* 31: doi.org/10.1016/j.tree.2016.01.012

952 Stamps, J.A. and A.M. Bell. 2020. The information provided by the absence of
953 cues: insights from Bayesian models of within and transgenerational plasticity. *Oecologia*
954 doi.org/10.1007/s00442-020-04792-9

955 Stiver, K.A. and S.H. Alonzo. 2013. Does the risk of sperm competition help ex-
956 plain cooperation between reproductive competitors? A study in the ocellated wrasse (
957 *Symphodus ocellatus*) *The American Naturalist* 81: 357-368

958 Tachiki, Y. and I. Koizumi. 2016. Absolute versus Relative Assessments of Individ-
959 ual Status in Status-Dependent Strategies in Stochastic Environments. *The American*
960 *Naturalist* 188: DOI: 10.1086/686899

961 Taper, M.L. and J. M. Ponciano. 2016. Evidential statistics as a statistical modern
962 synthesis to support 21st century science. *Population Ecology* 58:9-29

963 Thompson, J.N. 1986. Oviposition behaviour and searching efficiency in a natural
964 population of a braconid parasitoid. *Journal of Animal Ecology* 55:351-360

965 Thompson, J.N. 1987. Variance in the number of eggs per patch: oviposition behaviour
966 and population dispersion in a seed parasitic moth. *Ecological Entomology* 12:311-320

967 Thorpe, J.E., Mangel, M., Metcalfe, N.B., and F.A. Huntingford. 1998. Modelling
968 the proximate basis of life history variation, with application to Atlantic salmon, *Salmo*
969 *salar* L. *Evolutionary Ecology* 12:581-600

970 Trimmer, P.C. and Houston, A.I. 2014. An evolutionary perspective on information
971 processing. *Topics in Cognitive Science* 2014(2014):1-19

972 Trimmer, P.C., Ehlman, S.M, and A. Sih. 2017. Predicting behavioural responses to
973 novel organisms: state-dependent detection theory. *Proceedings of the Royal Society B*
974 284: 20162108. <http://dx.doi.org/10.1098/rspb.2016.2108>

975 Waddington, C. H. 1957. *The Strategy of Genes. A Discussion of Some Aspects of*
976 *Theoretical Biology*. George Allen & Unwin, London