

# The Evolution of Bayesian Updating

## Abstract

An evolutionary basis for Bayesian rationality is suggested, by considering how natural selection would operate on an organism's 'policy' for choosing an action depending on an environmental signal. It is shown that the evolutionarily optimal policy, as judged by the criterion of maximal expected reproductive output, is the policy which for each signal, chooses an action that maximizes *conditional* expected output given that signal. An organism using such a policy is behaving as if it were a Bayesian agent with probabilistic beliefs about the states of nature, that it updates by conditionalization, and whose choice behaviour obeys expected utility maximization. This suggests a possible route by which Bayes-rational creatures might have evolved. However this conclusion needs qualifying, since it relies on the assumption that expected reproductive output is the sole determinant of evolutionary success, which is not always true.

## 1 Introduction

Bayesian updating, also known as 'conditionalization', is a rule specifying how a prior probability distribution should be updated to a posterior distribution in the light of new information. The rule is often interpreted epistemically: it specifies how an agent should change their epistemic state over time in response to new evidence, where an 'epistemic state' is represented by a probability distribution over some specified set of alternatives.

There has been considerable philosophical discussion of the normative status of Bayesian updating. Does rationality require an agent to update using the Bayesian rule? If so, can the rule be justified from more primitive rationality requirements? Is it the unique rule that can be so justified? Are there situations where Bayesian updating should not be used? Works addressing these questions include Brown [1], Maher [20], van Fraassen [33], Christensen [2], Lewis [19], and more recently, Greaves and Wallace [13], and Leitgeb and Pettigrew [17].

Here I focus on a different though related issue. Instead of asking whether Bayesian updating is a requirement of *rationality*, I ask whether it is a requirement of *evolutionary optimality*. Changing one's beliefs in response to evidence is an aspect of cognition, and it seems likely that at least some aspects of cognition, animal and human, have been shaped by natural selection. This prompts the question of how selection would have operated on updating rules. An organism's updating rule affects its epistemic state, which in turn affects its behaviour;

so different updating rules will lead to different behaviours, and thus have different consequences for survival and reproduction. Can we show that organisms using Bayesian updating would have enjoyed a selective advantage over those updating in some other way?

This question is an aspect of the broader issue of whether natural selection will tend to produce rational behaviour. A number of authors have examined whether rational-choice norms such as transitivity of preference, maximisation of expected utility, and consistency of time-preference can be derived from an underlying evolutionary model (cf. Cooper [3], Houston, McNamara and Steer [15], Robson [26], Skyrms [30]). However the question of whether Bayesian updating can be derived from evolutionary principles has never been considered, so far as I know. Here I sketch a tentative answer to that question.

The structure of this paper is as follows. Section 2 briefly discusses the use of Bayesian concepts in evolutionary biology. Section 3 constructs a simple example to compare the fitness consequences of different ways that an organism's behaviour might be sensitive to information that it receives. Section 4 generalises the example into an abstract characterization of the optimal policy for choosing an action in the light of new information. Section 5 discusses how this optimal policy might be implemented by an organism. One possibility is that organism has an internal probabilistic representation of its environment, which it updates in a Bayesian manner; this suggest a possible evolutionary route to Bayesian updating. Section 6 relates our argument to certain 'pragmatic' arguments for Bayesian updating that have been made in a rational-choice context. Section 7 qualifies our argument by highlighting an implicit assumption on which it rests.

## 2 Bayesianism in Evolutionary Biology

Some readers may be surprised by our discussing Bayesian principles in an evolutionary context, on the grounds that non-human animals lack the cognitive sophistication required to have probabilistic beliefs about the world, less still to update them. But in fact, behavioural ecologists make widespread use of Bayesian ideas to think about animal behaviour, especially in foraging theory (cf. McNamara, Green and Olsson [23]).

A typical Bayesian approach in foraging theory is to assume that animals have some prior information about an environmental parameter, represented by a probability distribution. For example, the animal may 'know' that food patches are of two types, good and bad, whose relative frequencies are  $1/3$  and  $2/3$  respectively. This knowledge may either be genetically encoded or acquired by the animal through its experience. So before beginning foraging in any particular patch, the animal's prior belief that it is a good patch is  $1/3$ . As the animal begins to forage, it updates its belief about the type of patch it is in. If the probability that it is in a bad patch becomes sufficiently high, the animal may choose to move on. Many optimal foraging models aim to derive a precise prediction about when an animal should move from one patch to another.

In models of this sort, talk of an animal’s ‘knowledge’ or ‘information’ is usually intended behaviouristically. Saying that the animal ‘knows’ that  $1/3$  of food patches are good means that the animal behaves as if it knows this, as manifested in its choices. There is no great puzzle about how an animal could come to exhibit such behaviour. Natural selection leads to adaptive behaviour, one aspect of which is choosing the action appropriate to the environment. If the environment can be in a number of different states, and the appropriate action depends on the state, then to behave adaptively an animal will need to be sensitive to the states’ probabilities of occurrence. If the animal chooses between actions according to maximization of expected reproductive output – one standard definition of adaptive behaviour – then it is behaving as if it knows the true probability distribution on the states, knows the payoffs, and is able to calculate expectations.

Understanding ‘knowledge’ in this behaviouristic way may seem quite different from how the notion is understood in epistemology. But in fact many Bayesian decision theorists, notably Savage [28], have argued for a behaviouristic interpretation of probabilistic beliefs, according to which an agent’s subjective probabilities (and utilities) are derived from their choices between uncertain prospects, which are in principle observable; this interpretation is orthodox in contemporary economics, if not in contemporary epistemology. So in fact, evolutionary biologists’ use of Bayesian concepts to model animal behaviour is not such a radical departure.

There is abundant evidence that organisms of all taxa make adaptive use of information about their environment; see Giraldeau [10] for a review. In some cases natural selection can encode information about the environment into the genome, so organisms are born with innate knowledge of environmental parameters. In other cases selection cannot do this, as the environment changes too fast; but information about the environment can also be obtained by organisms during their lifetime, through experience. Often a combination of genetically-encoded and acquired information is used by organisms to guide their behaviour. These facts, which are almost platitudes in biology, do not imply that organisms are behaving like Bayesian agents; it is perfectly possible that they process information in a non-Bayesian way. But there is empirical evidence that organisms do sometimes reason in a Bayes-like manner, in that they appear to combine prior knowledge with new information to form an updated ‘worldview’, which then informs their behaviour<sup>1</sup> (cf. Valone [32]).

Interestingly, many facets of the theory of Bayesian rationality have evolutionary applications. For example, the famous ‘value of information’ theorem, due originally to F. P. Ramsey [25] and I. J. Good [11], has recently been applied in an evolutionary setting by McNamara and Dall [22]. The Ramsey/Good theorem states that cost-less information is always valuable, in that an agent who updates on new information before choosing an action will achieve expected utility no lower, and usually higher, than one choosing without the benefit of

---

<sup>1</sup>Again, this should be interpreted behaviouristically, i.e. the organisms behave as if they are incorporating the new information in a Bayesian way. It is not assumed that Bayesian calculations are actually going on in the organisms’ brains, though this is possible.

	No predator	snake	leopard
Stay put	10	0	0
Climb	5	4	0
Flee	6	1	2

Table 1: Payoffs for Three Alternative Actions

the information. McNamara and Dall [22] interpret the theorem biologically, as showing that information is a ‘fitness enhancing resource’, since it cannot reduce, and will usually increase, an organism’s expected reproductive output.

The ‘value of information’ theorem assumes that agents, or organisms, incorporate the new information by Bayesian updating. This assumption is crucial to the theorem; if non-Bayesian update rules are permitted, the desired result does not go through. Indeed the assumption of Bayesian updating is widespread in behavioural ecologists’ discussions of adaptive information use, but is never explicitly questioned. (This is reminiscent of the situation in probabilistic epistemology in the 1960s, where Bayesian updating was tacitly assumed but not explicitly discussed, until Hacking [14] observed that requiring an agent’s credences to satisfy the probability calculus at every instant does *not* imply that she use the Bayesian update rule.) In what follows, I explicitly consider the adaptive significance of Bayesian updating.

### 3 A simple example

To fix ideas, consider a simple example. An organism is foraging for food in a predator-strewn area. Predators are of two types, snakes and leopards. If no predator is present, the best thing to do is stay put and forage. If a snake is present, the best thing to do is to climb a tree. If a leopard is present, the best thing to do is flee. Both climbing and fleeing are costly in terms of time and energy, to different extents. Payoffs for the three actions, as a function of the state of the world, are shown in Table 1; these payoffs are measured in increments of biological fitness, i.e. number of offspring.

What action should the animal choose to maximise its expected payoff? That depends on the probabilities of the three states of the world. Suppose that the probabilities of the three states are:

$$P(\text{no predator}) = 1/2, P(\text{snake}) = 1/3, P(\text{leopard}) = 1/6$$

(These probabilities can be thought of as the relative frequencies with which each state occurs in the relevant local ecology.) Then, the expected payoff from each action is:

$$\begin{aligned} V(\text{stay put}) &= 10(1/2) + 0(1/3) + 0(1/6) = 5 \\ V(\text{climb}) &= 5(1/2) + 4(1/3) + 0(1/6) = 3.83 \\ V(\text{flee}) &= 6(1/2) + 1(1/3) + 2(1/6) = 3.67 \end{aligned}$$

So the evolutionarily optimal action is to stay put. Organisms choosing to stay put will on average leave more offspring than those choosing either of the other two actions. If the organism's choice behaviour has been optimised by natural selection, or if it has learnt which action is optimal, then it will stay put.

Now suppose that prior to choosing an action, the organism receives a signal which indicates whether a predator is present or not. So the signal has two values: 'safe' and 'unsafe'. The signal is perfectly reliable, i.e. indicates 'unsafe' iff a predator is present, but cannot discriminate between leopards and snakes. (So an unsafe signal means just that that the 'no predator' state of the world does not obtain.) The organism's choice of action may depend on which signal is received. So the organism needs to have a 'policy', i.e. a specification of which action to take for each value of the signal. There are 9 ( $=3^2$ ) possible policies, in this example. Let us consider three policies in particular.

Suppose firstly that the organism ignores the signal, and always chooses the action optimal for the situation in which no signal is received, i.e. stay put. So the organism's policy is 'if safe, stay put; if unsafe, stay put'. Let us denote this policy 'Ignore'. The policy may seem unpromising, intuitively, but it could reflect the organism's cognitive limitations. If an organism's choice behaviour has been fashioned by natural selection, but it is incapable of attending to the signal, or lacks behavioural plasticity, it might use a policy like 'Ignore'.

Secondly, suppose that the organism does attend to the signal, and adopts a 'maximin' strategy, i.e. it chooses the action which maximizes the minimum payoff it will receive, in the light of the information provided by the signal. So if the safe signal is received it chooses to stay put, obviously. If the unsafe signal is received it chooses to flee, as this guarantees it a payoff of at least 1. So its policy is 'if safe, stay put; if unsafe, flee'. Let us call this policy 'Maximin'. The policy reflects a high degree of risk aversion. By choosing to flee rather than climb when the unsafe signal is received, the organism forgoes a possible payoff of 4 in order to definitely avoid a payoff of 0.

Thirdly, suppose that the organism behaves like a Bayesian. On receipt of a signal, the organism chooses an action that maximises its *conditional* expected payoff, given the signal.<sup>2</sup> So if the safe signal is received, the organism chooses to stay put - as the conditional expected payoffs are then 10, 5 and 6 for stay put, climb and flee respectively. (This is because the conditional probability of the 'no predator' state, given the safe signal, is 1.) What if the unsafe signal is received? The conditional probabilities of the 'no predator', 'snake' and 'predator' states, given an unsafe signal, are 0, 2/3 and 1/3 respectively. The conditional expected payoffs for staying put, climbing and fleeing are then 0, 8/3 and 4/3 respectively, so the organism will choose to climb. Its policy is therefore 'if safe, stay put; if unsafe, climb'. Let us call this policy 'Bayes'.

Which of our three policies - 'Ignore', 'Maximin' and 'Bayes' - is the best, from an evolutionary point of view? To answer this question, we need to com-

---

<sup>2</sup>Note that this action need not be unique. If it is not unique, for one or more values of the signal, then 'the Bayes policy' is really a class of policies each of which achieves maximum conditional expected payoff. See section 4.

pute the expected payoffs accruing to an organism that uses each of the policies. Consider firstly ‘Ignore’. Since an organism using ‘Ignore’ chooses to stay put whether or not a signal is received, its expected payoff is:

$$V [\text{Ignore}] = 10(1/2) + 0(1/3) + 0(1/6) = 5$$

What about ‘Maximin’? An organism using ‘Maximin’ stays put if a safe signal is received but flees otherwise. With probability 1/2 no predator is present, so the safe signal is sent, so the organism stays put and earns a payoff of 10. With probability 1/3 a snake is present, so the unsafe signal is sent, so the organism flees and earns a payoff of 1. With probability 1/6 a leopard is present, so the unsafe signal is sent, so the organism flees and earns a payoff of 2. Its expected payoff is therefore

$$V [\text{Maximin}] = 10(1/2) + 1(1/3) + 2(1/6) = 5.67$$

By a similar logic, the payoff to an organism using the Bayes policy is :

$$V [\text{Bayes}] = 10(1/2) + 4(1/3) + 0(1/6) = 6.33$$

Therefore, the Bayes policy yields the highest expected payoff. If all three policies are found in a population, natural selection will favour the Bayes policy over the other two. Over time, evolution should convert the population to the Bayes policy, driving the other two policies extinct. This suggests, obviously in a preliminary way, that evolution might produce organisms that behave as if they were rational Bayesian agents, whose subjective probabilities over the states of nature match the objective frequencies and who incorporate new information by conditionalization. A stronger argument for this conclusion would need to show that no conceivable policy does better than the Bayes policy; see section 4.

Both the Bayes and the Maximin policies, in this example, may seem computationally demanding. The Bayes policy is defined as that policy which for each signal, picks an action that maximizes conditional expected payoff given that signal. However we need not assume that an organism, to implement the Bayes policy, actually computes conditional expected payoffs. If the organism, on receipt of a signal, chooses an action that does *in fact* maximize conditional expected payoff, then it is by definition implementing the Bayes policy. A ‘policy’, as we have defined it, is simply a complex behavioural disposition, or function from signals to actions. How an organism might implement a policy is discussed in section 5.

Our simple example bears a close relation to the biological version of the Ramsey/Good ‘value of information’ theorem. In effect, the Ramsey/Good theorem shows that an organism using the Bayes policy will achieve greater expected payoff, hence evolutionary success, than the policy we have called ‘Ignore’. But it says nothing about other non-Bayesian policies, such as ‘Maximin’ for example. However, it is straightforward to show that the Bayes policy outperforms any other policy, so is evolutionarily optimal. That is the task of the next section.

## 4 Can't do Better than Bayes <sup>3</sup>

The framework we adopt directly generalizes the previous example. There is a finite set  $S$  of states of nature;  $S = \{\theta_1, \dots, \theta_k\}$ . There is a finite set  $U$  of actions that the organism may perform. The payoff from an action depends on the state of nature. The payoff from action  $u \in U$  if the state of nature turns out to be  $\theta_i$  is  $V_i(u)$ . As before, payoffs are measured in increments of biological fitness.

There is a finite set  $E$  of possible signals, or items of evidence, that the organism may receive;  $E = \{E_1, \dots, E_n\}$ . In the example above we assumed that the signal set partitioned the states of nature, i.e. each state of nature was compatible with exactly one signal, but no such assumption is made here. There is a joint probability distribution  $p$  on  $S \times E$ , reflecting the frequency with which each (state, signal) pair occurs. The marginal probability of state  $\theta_i$  is  $p(\theta_i)$ , and of signal  $E_j$  is  $p(E_j)$ . We assume that  $p(\theta_i) > 0$  for all  $i$  and  $p(E_j) > 0$  for all  $j$ , i.e. each state occurs with positive probability, and similarly for each signal.

An organism's 'policy' specifies an action  $u \in U$  for every possible signal  $E_j \in E$ . Thus a policy is simply a function from  $E$  to  $U$ . The set of all policies is denoted  $X$ . For any policy  $x \in X$ , we let  $x(E_j) \in U$  be the action specified by that policy when signal  $E_j$  is received; for convenience, we will write  $x(E_j)$  as  $x_j$ .

Consider an arbitrary policy  $x \in X$ . What is an organism's expected payoff from using policy  $x$ ? Suppose firstly that the true state of nature is  $\theta_i$ . Then, the expected payoff from policy  $x$  equals:

$$\sum_{E_j} p(E_j/\theta_i).V_i(x_j) \tag{1}$$

The justification for (1) is clear.  $p(E_j/\theta_i)$  is the probability that signal  $E_j$  is received given that the state of nature is  $\theta_i$ .  $V_i(x_j)$  is the payoff from choosing action  $x_j$  – the action specified by policy  $x$  when the signal received is  $E_j$  – in state of nature  $\theta_i$ . Expression (1) is thus the expected payoff to an organism using policy  $x$  when the true state is  $\theta_i$ . Taking the expectation of (1) across states of nature thus gives us the overall **expected payoff to policy  $x$** :

$$\sum_{\theta_i} p(\theta_i) \sum_{E_j} p(E_j/\theta_i).V_i(x_j) \tag{2}$$

Now recall the Bayes policy discussed above. In our simple example above, there was a unique Bayes policy, since for each signal there was a unique action that maximized conditional expected payoff given the signal. But in the general case this need not be so: for some signals, there may be more than one action satisfying this maximization constraint. So we need to consider the class of Bayes policies  $B \subset X$ , where each  $b \in B$  is a policy that, for each signal  $E_j \in E$  chooses an action  $b_j \in U$  that maximises conditional expected payoff given  $E_j$ .

---

<sup>3</sup>Thanks to John McNamara and Cedric Paternotte for help with this section.

We wish to show that a policy is evolutionarily optimal if and only if it is a Bayes policy; as before, ‘evolutionarily optimal’ means ‘earns an expected payoff greater than or equal to every other policy’. To show this, consider any Bayes policy  $b \in B$ . Substituting  $b$  for  $x$  in expression (2) thus gives us the **expected payoff to a Bayes policy** :

$$\sum_{\theta_i} p(\theta_i) \sum_{E_j} p(E_j/\theta_i) \cdot V_i(b_j) \quad (3)$$

Our first task is to show that expression (3) is greater than or equal to (2), i.e. policy  $b$  has expected payoff greater than or equal to policy  $x$ .

Since  $b$  is a Bayes policy, we know that it satisfies the condition:

For every signal  $E_j$ ,

$$\sum_{\theta_i} p(\theta_i/E_j) \cdot V_i(b_j) \geq \sum_{\theta_i} p(\theta_i/E_j) \cdot V_i(u) \text{ for all actions } u \in U \quad (4)$$

Applying inequality (4) to the particular action  $x_j \in U$  gives:

$$\sum_{\theta_i} p(\theta_i/E_j) \cdot V_i(b_j) \geq \sum_{\theta_i} p(\theta_i/E_j) \cdot V_i(x_j) \quad (5)$$

Applying Bayes’ theorem and dividing across by  $p(E_j)$ :

$$\sum_{\theta_i} p(\theta_i) \cdot p(E_j/\theta_i) \cdot V_i(b_j) \geq \sum_{\theta_i} p(\theta_i) \cdot p(E_j/\theta_i) \cdot V_i(x_j) \quad (6)$$

Summing over all the signals:

$$\sum_{E_j} \sum_{\theta_i} p(\theta_i) \cdot p(E_j/\theta_i) \cdot V_i(b_j) \geq \sum_{E_j} \sum_{\theta_i} p(\theta_i) \cdot p(E_j/\theta_i) \cdot V_i(x_j) \quad (7)$$

Reversing the order of summation:

$$\sum_{\theta_i} p(\theta_i) \sum_{E_j} p(E_j/\theta_i) \cdot V_i(b_j) \geq \sum_{\theta_i} p(\theta_i) \sum_{E_j} p(E_j/\theta_i) \cdot V_i(x_j) \quad (8)$$

But the LHS of (8) is the expected payoff to policy  $b$  (expression (3)), while the RHS is the expected payoff to policy  $x$  (expression (2)). So policy  $b$  is evolutionary optimal. Since nothing has been assumed about policy  $b$  except that it is a Bayes policy, we can conclude that every Bayes policy is evolutionarily optimal.

To show the converse, consider another policy  $y$  which is *not* a Bayes policy. Thus for some signal  $E_c$ :

$$\sum_{\theta_i} p(\theta_i/E_c) \cdot V_i(y_c) < \sum_{\theta_i} p(\theta_i/E_c) \cdot V_i(b_c) \quad (9)$$

Applying Bayes’ theorem:



$$\sum_{\theta_i} p(\theta_i) \cdot p(E_c/\theta_i) \cdot V_i(y_c) < \sum_{\theta_i} p(\theta_i) \cdot p(E_c/\theta_i) \cdot V_i(b_c) \quad (10)$$

Since  $b$  is a Bayes policy, we know from (6) that for all signals  $E_j$ :

$$\sum_{\theta_i} p(\theta_i) \cdot p(E_j/\theta_i) \cdot V_i(y_j) \leq \sum_{\theta_i} p(\theta_i) \cdot p(E_j/\theta_i) \cdot V_i(b_j) \quad (11)$$

From (10) and (11), we can deduce that, summing over all the signals:

$$\sum_{E_j} \sum_{\theta_i} p(\theta_i) \cdot p(E_j/\theta_i) \cdot V_i(y_j) < \sum_{E_j} \sum_{\theta_i} p(\theta_i) \cdot p(E_j/\theta_i) \cdot V_i(b_j) \quad (12)$$

Reversing the order of summation:

$$\sum_{\theta_i} p(\theta_i) \sum_{E_j} p(E_j/\theta_i) \cdot V_i(y_j) < \sum_{\theta_i} p(\theta_i) \sum_{E_j} p(E_j/\theta_i) \cdot V_i(b_j) \quad (13)$$

But the LHS of (13) is the expected payoff to policy  $y$ , while the RHS is the expected payoff to the Bayes policy  $b$ . So any policy that is *not* a Bayes policy earns a strictly lower expected payoff than a Bayes policy, so cannot be evolutionarily optimal. Therefore an evolutionarily optimal policy must be a Bayes policy.

Taken together, (8) and (13) tell us that a policy is evolutionarily optimal if and only if it is a Bayes policy. So organisms that implement Bayes policies will leave more offspring, on average, than ones that do not. Over time, we would expect the non-Bayes policies to be eliminated by natural selection.

## 5 Discussion

What exactly does the foregoing argument show? The argument *seems* to supply an evolutionary basis for Bayesian updating, by showing that it follows from the standard Darwinian assumption that animal behaviour is well-adapted, hence designed to maximise the animal's expected reproductive output. However, some care is needed before adopting this interpretation. Typical discussions of Bayesian updating ask how an agent does or should update their epistemic state on receipt of information, where an epistemic state is represented by a prior probability distribution over some appropriate algebra. But our argument above makes no mention of 'epistemic states' or ways of updating them, and makes no assumption that the organisms in question are capable of being in such states. We have assumed only that the organisms are behaviourally plastic, and capable of modifying their behaviour in response to an environmental signal. Our argument provides a precise characterization of the 'best' way for an organism to let its choice of action depend on the signal it receives, as judged by the criterion of maximum expected reproductive output.

To get from this behaviourist starting point to something more epistemic, we need to consider how an organism might implement an evolutionarily optimal policy, as characterized above. One possibility is this: the organism operates with an internal ‘belief-like’ representation of the environment, an internal ‘desire-like’ representation of the outcomes (or payoffs), and uses a ‘choice rule’ to determine which action to pick, depending on which belief and desire states it is in. On receipt of a signal, the organism uses an ‘update rule’ to go from one belief state to another. Given a generic psychological make-up of this sort, which is presumably applicable to at least some higher animals as well as humans, an evolutionarily optimal policy is straightforward to implement.

To see this, suppose we take a ‘belief state’ to be a subjective probability distribution over the states of nature and signals, and a ‘desire state’ to be a real-valued utility function on the outcomes (or action-state pairs). In particular, suppose that the organism’s initial belief state, prior to receiving a signal, is equal to the true prior distribution  $p$ , and its utility function is equal to the true payoff function  $V$  (or an affine transformation thereof). Suppose that the organism’s choice rule is ‘maximize expected utility, relative to my current belief state’, and that its update rule is Bayesian conditionalization. Then, it follows that the organism will implement an evolutionarily optimal policy as characterized in section 4; this can be seen by direct inspection of expression (3).

What could license the assumptions that the organism’s initial belief state is correct, i.e. matches the true distribution on the states, and that its utility function matches the true payoff function? A short answer is ‘evolution by natural selection’. As discussed in section 2, organisms often have information about the environment ‘pre-programmed’ into their genome, as a result of natural selection in the past, and are often equipped with ‘appropriate’ desires for surviving and reproducing. (For example many animals are born with an innate ability to determine who their close relatives are, and innate preferences for some food items over others.) In short, it is plausibly an adaptive advantage to have correct probabilistic beliefs about the world, and desires for things that objectively enhance reproductive success. From an adaptationist perspective, therefore, the assumptions in question do not seem unreasonable.

Given these assumptions, our argument shows that an organism will implement an evolutionarily optimal policy if it chooses between actions in accordance with expected utility maximization relative to its current belief state, and always updates its belief state by the rule of conditionalization. Therefore, two standard principles of Bayesian rationality fall out of the requirement of evolutionary optimality. This suggests that, other things being equal, natural selection should lead to organisms that satisfy these Bayesian principles.

This conclusion needs qualifying in a number of ways. Firstly, the inference from ‘is evolutionarily optimal’ to ‘is likely to have evolved’ can obviously be questioned; this is a well-known issue in the literature on adaptationism. Secondly, defining optimality in terms of maximal expected reproductive output raises some tricky theoretical issues, discussed in the next section. Thirdly, our argument does not show that the *only* way for an organism to implement an

evolutionarily optimal policy is to operate with internal belief-like and desire-like representations which combine and update in accordance with Bayesian principles. This last point merits some brief discussion.

It is well-known from the work of Gerd Gigerenzer and colleagues [7, 8] that adaptive behaviour may be produced by simple heuristics, or rules-of-thumb, which are computationally less demanding than optimality calculations. This general point applies to our problem of choosing an optimal policy, or function from signals to actions. Even if there are many states of nature and many signals, an organism might still be able to implement an optimal policy by a simple heuristic such as ‘if signal  $x$  is received, run away; otherwise, stay’, depending on the payoff function. And in fact, many simple animals exhibit behavioural plasticity in response to environmental cues but presumably do not operate with a belief-like representation of the environment at all; so the issue of how to update does not arise. Of course in a purely ‘as if’ sense, these animals are behaving like Bayesian agents so long as they are implementing an optimal policy, as our argument shows. But we should not blur the distinction between organisms which implement an optimal policy by using a heuristic, and ones which operate with internal belief-like and desire-like representations and are thus capable of satisfying Bayesian principles in more than an ‘as if’ sense. An important question is when evolution will lead to organisms of each type. This question is not addressed by the foregoing analysis; see Sterelny [31] for a sketch of a possible answer.

To sum up: one way an organism might implement an evolutionarily optimal policy, though not the only way, is to follow standard Bayesian maxims for choosing between actions and for updating beliefs. So long as the organism’s initial belief state and utility function are ‘objectively correct’ in the sense specified above, adhering to these Bayesian maxims is sufficient to ensure that the organism’s policy, i.e. function from signals to actions, is evolutionarily optimal. This suggests a possible route by which Bayes-rational creatures might have evolved.

## 6 Relation to the Brown/Maher argument

It is worth relating our evolutionary argument to a related argument found in the rational choice literature. Both Brown [1] and Maher [20] try to justify Bayesian conditionalization by arguing that it follows from the injunction to maximize expected utility.<sup>4</sup> They consider an agent with a prior subjective probability distribution over a set of states of nature, a set of possible actions, and a utility function defined on the (state, action) pairs. The agent receives a signal, updates to a new probability distribution, then chooses an action that maximises expected utility relative to this new distribution. Brown and Maher then show that in order to maximize expected payoff with respect to the agent’s

---

<sup>4</sup>A related argument is given by Greaves and Wallace [13]. All of these arguments, and our own, are arguably just elaborations on a point first made by Ramsey [25].

prior distribution, their update rule should be conditionalization. They interpret this to mean that rationality requires an agent to update in a Bayesian way.

What is the logical relation between the Brown/Maher result and our own? Our result shows that, to maximise expected payoff relative to a prior distribution  $p$ , an organism should choose a policy which, for each signal, maximizes the conditional expected payoff given the signal. The Brown/Maher result shows that, to maximize expected payoff relative to a prior distribution  $p$ , an agent should update by conditionalization, *given that its choice of action in its updated state goes by maximization of expected payoff relative to that state*. In our framework the prior  $p$  is interpreted as the objective frequency distribution on the states and the payoff function as the fitness function; while in the Brown/Maher framework, the prior  $p$  is the agent's subjective prior and the payoff function is their utility function. But leaving aside these interpretive differences, what is the formal relationship between the results?

The answer is that the Brown/Maher result is a special case of our own.<sup>5</sup> In effect, the Brown/Maher argument identifies a particular way of implementing an optimal policy, for an agent who uses an update rule to go from belief state to belief state, and a choice rule to go from belief state and utility function to choice of action. Brown/Maher assume that the choice rule is expected utility maximization relative to the updated state, and then show that the update rule should be conditionalization. Our result is more general, in that it does not assume from the outset that the organism implements its policy by using an (update rule, choice rule) combination at all. However, it follows directly from our result that *if* an organism implements its policy this way, and *if* its choice rule is 'maximize expected payoff relative to the updated state', then its update rule should be conditionalization – which is the Brown/Maher result. This follows directly from the characterization of an optimal policy in expression (3).

To see that our result is stronger than the Brown/Maher result, notice that their's leaves open the possibility that an agent could earn a higher expected payoff, relative to the prior distribution  $p$ , by using an update rule other than conditionalization and a choice-of-action rule other than maximize expected payoff. This possibility is not ruled out by the Brown/Maher argument, since they take for granted that the choice-of-action rule is maximize expected payoff. (This is quite reasonable given their aim. It would be rather odd for a rational agent to use expected payoff maximization in the prior state  $p$  to decide what update rule to employ, and yet *not* use expected payoff maximization in the updated state to decide what action to choose. But it is a logical possibility.) However, our result rules the possibility in question out. Since the combination (conditionalization, expected payoff maximization) implements an optimal policy, it follows that no other combination can do better, even if choice-of-action

---

<sup>5</sup>Two slight qualifications are needed here. Brown's argument makes the restrictive assumption that the signal set partitions the set of states of nature, as in our simple example in section 3, which is a special case. Maher's argument relaxes this assumption and also adopts the framework of 'causal decision theory' in the sense of Lewis [18]. Our formal framework of section 4 also relaxes the partitioning assumption, but does not deal with the additional complexities raised by causal decision theory.

rules other than expected payoff maximization are permitted.

## 7 A Qualification

The foregoing argument rests on an important assumption, namely that expected payoff, i.e. expected reproductive output, is the right criterion of evolutionary success. But in fact this is not always the case. It is well-known that in certain situations, the expected performance of a strategy (or phenotype) is not the sole determinant of whether it will evolve; variability in performance can also matter (cf. Frank and Slatkin [6], Gillespie [9], Seger and Brockmann [29]). How does this complication affect the foregoing argument?

Expected reproductive output is an appropriate criterion of evolutionary success (or ‘definition of fitness’) when two conditions are met. Firstly the population must be large, and secondly the probability distribution over states of nature that any organism faces must be independent across the members of the population. These conditions jointly ensure that the average reproductive success of the cohort of organisms using a given strategy will be very close to that strategy’s expected success. The large population condition will often be satisfied, but the independence condition is more problematic.

In the predator example of section 3, the independence condition seems reasonable. The probability that organism A encounters a snake while foraging may well be independent of the probability that organism B does; risks of this sort are sometimes called ‘idiosyncratic’. But in other cases risks will be aggregate. The weather is an obvious example, as it affects many organisms at once. If there is a 5% chance of a harsh winter, this risk will obviously not be independent across all members of a biological population. Real organisms probably face both idiosyncratic and aggregate risks, in varying combinations depending on the context.<sup>6</sup>

Where risks are aggregate, the strategy with the highest expected reproductive output need not be evolutionarily optimal. The variance in output matters too, and selection will penalize strategies with a high variance. In such a case, our simple argument for the evolutionary optimality of Bayesian updating does not apply.

Robson [27] has shown that where there is a component of aggregate risk, ‘irrational’ behaviour may evolve, in that organisms whose choice behaviour violates expected utility maximization may have a selective advantage; a similar result was independently shown by McNamara [21]. This remarkable result arises because the determinant of evolutionary success is how well one does *relative to others*. With aggregate risk, optimal behaviour requires that an organism use a biased probability distribution, which shifts probability mass away from states of nature where the whole population does well and onto states where it does badly. The optimal behaviour is the one that maximises ‘expected’

---

<sup>6</sup>Note that idiosyncratic and aggregate are the ends of a continuum; intermediate degrees of correlation are also possible.

payoff relative to this biased distribution, rather than the true distribution.<sup>7</sup> It follows that evolution will favour choice behaviour that is sensitive to the variance as well as the expectation of reproductive output, and thus that violates expected utility maximization.

A precisely parallel argument could be constructed for the case of Bayesian updating. With aggregate risk, an organism which uses a non-Bayesian updating rule could conceivably have a selective advantage over one which uses Bayesian updating, so ‘irrational’ updating could evolve. This follows directly from the Robson/McNamara argument for how expected utility violations can evolve, combined with the argument of the previous section. If an organism implements its policy, i.e. function from signals to actions, by using an (update rule, choice-of-action rule) combination, and if its choice-of-action rule is ‘maximize expected payoff relative to the updated belief state’, then conditionalization will be the optimal update rule if and only if expected payoff maximization is the appropriate criterion of optimality.

This suggests that in theory, organisms whose behaviour is mediated by internal belief-like and desire-like states might do best to use Bayesian updating some of the time but not always. In a circumstance where risks are idiosyncratic, as in our predator example, an organism should incorporate new information by Bayesian updating, assuming that its choice-of-action rule is ‘maximize expected payoff in the updated state’. But where risks are aggregate, as in the weather example, a non-Bayesian update rule may be evolutionarily superior. (Nothing general can be said about what the best non-Bayesian rule will be; it depends on the details of the example.) However to implement this selectively Bayesian strategy would be extremely difficult, for it would require being able to distinguish aggregate from idiosyncratic risks. It seems unlikely that organisms can do this.

An interesting response to the Robson/McNamara argument comes from Grafen [12] and Curry [4], who argue that rational behaviour can be restored if in every state of nature, payoffs are computed relative to the population average; see Okasha [24] for discussion. So an organism needs to consider not the absolute number of offspring that an action will bring in a given state of nature, but rather what fraction of the total population’s reproduction it will achieve in that state. The expected value of this fraction across states of nature provides a correct criterion of evolutionary success, even where risks are aggregate. In short, evolutionary optimality requires maximising expected *relative* reproductive output. If an organism’s utility function depends suitably on relative reproductive output, expected utility maximisation will be evolutionarily optimal.

The Curry/Grafen point implies that in principle, our argument for the evolution of Bayesian updating can be rescued. However again, it seems very unlikely that an organism can know the relative reproductive output that an action will bring – for this depends on the rest of the population. Tailoring its

---

<sup>7</sup>Conceptually this is similar to what decision theorists call non-expected utility maximisation; see Okasha [24] for discussion.

choices to the criterion of expected relative output will probably be cognitively impossible. Most likely, the best the organism can do is to use absolute reproductive output as a proxy for relative output, and attempt to maximise the former. But this is an empirical claim; it would be falsified if it were discovered that animals do in fact respond differently to aggregate and idiosyncratic risks.

Where does this leave us? The criterion of expected reproductive output is commonly used by evolutionary modellers, even though it is known not to be universally applicable. Where it is inapplicable, our argument for the evolutionary optimality of Bayesian updating does not work. In principle the argument can be salvaged by the Curry/Grafen move, but at the expense of making implausible assumptions about the ability of organisms to tailor their choices to relative reproductive success. So on balance, there is certainly some reason to think that Bayesian updating will evolve by natural selection, but the case is not watertight.

A final consideration is this. It is well-known that humans are not good Bayesians, in either a conscious or an ‘as if’ sense. Even intelligent people are notoriously poor at explicitly calculating conditional probabilities, and human choice behaviour is typically inconsistent with our having underlying probabilistic beliefs at all (Kahneman, Tversky and Slovic [16], Ellsberg [5]). If Bayesian updating is adaptively advantageous, it might seem surprising that humans are not better at it. Seen in this light, the fact that our evolutionary argument is not watertight is perhaps no bad thing. This is not to say that the reason why our argument is not watertight – the fact that some risks are aggregate – explains why humans have not evolved to be good Bayesians. But it is an intriguing possibility, worthy of further exploration.

## References

- [1] P. M. Brown. Discussion: Conditionalization and expected utility. *Philosophy of Science*, 43:415–419, 1976.
- [2] D. Christensen. Clever bookies and coherent beliefs. *Philosophical Review*, (2):229–247, 1991.
- [3] W. S. Cooper. *The Evolution of Reason*. Cambridge University Press, Cambridge, 2003.
- [4] P. Curry. Decision making under uncertainty and the evolution of interdependent preferences. *Journal of Economic Theory*, 98:357–369, 2001.
- [5] D. Ellsberg. Risk, ambiguity and the Ellsberg paradox. *Quarterly Journal of Economics*, 75(4):643–669, 1961.
- [6] S. A. Frank and M. Slatkin. Evolution in variable environments. *American Naturalist*, 136(2):244–260, 1990.
- [7] G. Gigerenzer and P. Todd. *Simple Heuristics that Make us Smart*. Oxford University Press, Oxford, 1999.

- [8] G. Gigerenzer, P. Todd, and T. Pachur, editors. *Heuristics: the Foundations of Adaptive Behaviour*. Oxford University Press, Oxford, 2011.
- [9] J. Gillespie. Natural selection for variances in offspring number: a new evolutionary principle. *American Naturalist*, 111:1010–1014, 1977.
- [10] L. Giraldeau. The ecology of information use, 4th edition. In *Behavioural Ecology*, pages 42–88. Blackwell, Oxford, 1997.
- [11] I. J. Good. On the principle of total evidence. *British Journal for the Philosophy of Science*, 17:319–321, 1967.
- [12] A. Grafen. Formal darwinism, the individual-as-maximising-agent analogy, and bet-hedging. *Proceedings of the Royal Society B*, 266:799–803, 1999.
- [13] H. Greaves and D. Wallace. Justifying conditionalization: Conditionalization maximizes expected epistemic utility. *Mind*, 115:607–632, 2006.
- [14] I. Hacking. Slightly more realistic personal probability. *Philosophy of Science*, 34(4):311–325, 1967.
- [15] A. I. Houston, J. M. McNamara, and M. D. Steer. Do we expect natural selection to produce rational behaviour? *Philosophical Transactions of the Royal Society B*, 362:1531–1543, 2007.
- [16] D. Kahneman, P. Slovic, and A. Tversky, editors. *Judgement under Uncertainty Heuristics and Biases*. Cambridge University Press, Oxford, 1982.
- [17] H. Leitgeb and R. Pettigrew. An objective justification of Bayesianism II: the consequences of minimizing inaccuracy. *Philosophy of Science*, 77:236–272, 2010.
- [18] D. Lewis. Causal decision theory. *Australasian Journal of Philosophy*, 59:5–30, 1981.
- [19] D. Lewis. Why conditionalize? In *Papers in Metaphysics and Epistemology*, pages 403–407. Cambridge University Press, Cambridge, 1999.
- [20] P. Maher. Diachronic rationality. *Philosophy of Science*, 59:120–141, 1992.
- [21] J. M. McNamara. Implicit frequency-dependence and kin selection in fluctuating environments. *Evolutionary Ecology*, 9:185–203, 1995.
- [22] J. M. McNamara and S. Dall. Information is a fitness-enhancing resource. *Oikos*, 119:231–236, 2010.
- [23] J. M. McNamara, R. F. Green, and O. Olsson. Bayes’ theorem and its applications in animal behaviour. *Oikos*, 112:243–251, 2006.
- [24] S. Okasha. Optional choice in the face of risk: Decision theory meets evolution. *Philosophy of Science*, forthcoming 2011.



- [25] F. P. Ramsey. Weight or the value of knowledge. *British Journal for the Philosophy of Science*, 41:1–4, 1990.
- [26] A. Robson. Evolution and human nature. *Journal of Economic Perspectives*, 16:89–106.
- [27] A. Robson. A biological basis for expected and non-expected utility. *Journal of Economic Theory*, 68:397–424, 1996.
- [28] L. J. Savage. *The Foundations of Statistics*. Wiley, New York, 1954.
- [29] J. Seger and H. J. Brockmann. What is bet-hedging? *Oxford Surveys in Evolutionary Biology*, 4:182–211, 2002.
- [30] B. Skyrms. *Evolution of the Social Contract*. Cambridge University Press, Cambridge, 1995.
- [31] K. Sterelny. *Thought in a Hostile World*. Blackwell, Oxford, 2003.
- [32] T. J. Valone. Are animals capable of Bayesian updating? *Oikos*, 112:252–259, 2006.
- [33] B. van Fraassen. *Laws and Symmetry*. Oxford University Press, Oxford, 1989.