Learning and prejudice

The adaptive advantage of learning and *a priori* prejudice

1. INTRODUCTION

Suppose an organism sees that a tiger is at hand. The organism must decide whether *this tiger is dangerous*. Two types of strategy are available for making this decision.

An individual who *learns* will believe the proposition, or believe its negation, depending on the character of its experiences. For example, the organism might attend to whether the tiger is wagging its tail, and decide what to believe on that basis. The alternative strategy is for the organism to decide on the basis of *a priori prejudice*. It will believe that the tiger is dangerous (or that it is not) irrespective of the character of its experience.

Learning is a conditional strategy, whereas adhering to an *a priori* prejudice is unconditional. The prejudiced individual conforms to the rule *always believe that tigers are dangerous*.

The learner conforms to the rule believe that a tiger is dangerous if your experience has characteristic C, but believe that the tiger is not dangerous if your experience has characteristic D.

The problem to be investigated here concerns the adaptive advantage of learning and *a priori* prejudice. Under what circumstances is learning advantageous? This problem is subsumable under the more general heading of determining when an *obligate response* is fitter than a *facultative response*. A polar bear can either have thick fur regardless of what the ambient temperature is, or it can have thick fur in some circum-


stances and thin fur in others. Which strategy would be better for the polar bear to follow?

When I speak of an organism "deciding *a priori*" to believe that a tiger is dangerous, I do not mean that the organism is making a decision on its own. Rather, evolution has wired the organism to believe this proposition unconditionally. For exploratory purposes, I will adopt the working hypothesis that the traits in question are shaped by natural selection. So our question really is: When will natural selection favor learning over *a priori* prejudice?

In seeking an answer, we should not assume that the best policy with respect to one proposition will automatically be the best policy with respect to another. Maybe some problems are best approached by learning, while others are best solved *a priori*. And perhaps some organisms approach a given proposition via learning while others settle the same matter *a priori*. What is an *a priori* prejudice for one species may be a product of learning for another.

2. THE MODEL

Whether learning is better than prejudice depends on a variety of factors, which we now must try to identify. We begin by specifying the costs and benefits imposed by true and false belief:

\[
\begin{array}{c|c|c}
A \text{ is true} & A \text{ is false} \\
\hline
B e l i e v e \; A & x + b_1 & x - c_1 \\
B e l i e v e \; n o t-A & x - c_1 & x + b_2 \\
\end{array}
\]

It is not an assumption of this model that true belief is superior to falsehood. In what follows we will consider this idea (that \(x + b_1 > x - c_1\), \(i = 1,2\)) as one possibility among several.

We also must consider the probability that the agent who learns will form a given belief, conditional on the way the world is:

\[
\begin{array}{c|c|c}
A \text{ is true} & A \text{ is false} \\
\hline
B e l i e v e \; A & 1 - a & n \\
B e l i e v e \; n o t-A & a & 1 - n \\
\end{array}
\]

In this model, \(a\) and \(n\) are error probabilities; low values for \(a\) and \(n\) mean that if \(A\) is true (false), the agent will probably come to believe that \(A\) is true (false).

I will depart somewhat from scientific usage by saying that \(a\) and \(n\)
measure how *sensitive* the organism is. They describe how probable it is that the organism will believe a given proposition, if that proposition is true. Sensitivity, so defined, is a quite different matter from reliability. Reliability means that a proposition is probably true, if the agent believes that it is. As I use the terms, sensitivity is a world-to-head relation, while reliability is a head-to-world relation. It is not hard to see how the concepts differ. An agent who believes only those propositions that are absolutely certain will be quite reliable, though very insensitive. In any event, our model, as described so far, describes sensitivity, not reliability, though we will come to reliability in a while.

We now can describe three strategies. First, there are two forms of prejudice to consider. An organism can unconditionally believe that \( A \) is true or unconditionally believe that \( A \) is false. The third strategy is to decide whether to believe \( A \) or not-\( A \) based on the learning mechanism whose error characteristics were just described. Where \( p \) is the probability that \( A \) is true, the fitnesses of the three strategies are

\[
\begin{align*}
    w_{\text{Prej}(A)} &= (x + b_1)p + (x - c_2)(1 - p) \\
    w_{\text{Prej}(\neg A)} &= (x - c_1)p + (x + b_2)(1 - p) \\
    w_{\text{Learn}} &= [(x + b_1)(1 - a) + (x - c_1)a]p \\
    &+ [(x + b_2)(1 - n) + (x - c_2)n](1 - p)
\end{align*}
\]

Two inequalities follow:

1. \( w_{\text{Learn}} > w_{\text{Prej}(A)} \) if and only if \( (b_1 + c_2)(1 - n)(1 - p) > (b_1 + c_1)p \).
2. \( w_{\text{Learn}} > w_{\text{Prej}(\neg A)} \) if and only if \( (b_1 + c_1)p(1 - a) > (b_2 + c_2)(1 - p)n \).

3. A PROPOSITION'S IMPORTANCE

Exploring the implications of this model requires that we define a new concept:

the importance of a proposition \( A_i = a_i(b_i + c_i) \).

The importance of a proposition \( A_i \) is measured by seeing how much difference it makes whether one believes \( A_i \) or believes not-\( A_i \), if \( A_i \) is true; \( (x + b_1) - (x - c_2) = b_1 + c_2 \).

This definition induces a threefold division among propositions: a proposition \( A_i \) can have positive importance \( (b_i + c_i > 0) \), negative importance \( (b_i + c_i < 0) \), or zero importance \( (b_i + c_i = 0) \). Examples of the first type are familiar: If a mushroom is poisonous, it may be better to believe this than to believe its negation. For examples of the third kind, the reader need look no farther than the propositions of the present essay. I suspect that they are of no importance; even if they are true, it makes no difference in fitness whether you believe them or their negations.

The second sort of proposition is the least familiar of the three, but it is by no means impossible. A contemporary of Darwin's once opined that it would be better not to believe the theory of evolution even if it happens to be true (Ruse 1979). The same has been said more recently of utilitarianism (Williams 1972). And on a more personal level, the thought has been formulated that we should believe that life has a meaning, even if it does not. Of course, I am not endorsing any of these claims; I mention them just to illustrate what it means to say that a proposition has negative importance.

It is plausible that some propositions and their negations will both have positive importance. For example, in the pair of propositions (A tiger is nearby, No tiger is nearby), it may be better to believe the truth, whatever that happens to be. But in other examples, a different pattern may obtain. For example, in "The Will to Believe" William James suggests that it is better to believe in God, regardless of whether God in fact exists. One receives comfort either way, and that is the main thing (James 1897). This amounts to saying that the proposition "God exists" has positive importance and that "God does not exist" has negative importance. Again my point is not to endorse the claim, but to illustrate how a proposition and its negation need not have the same type of importance.

4. SOME CONSEQUENCES

We may begin with the question of how the relationship of learning to a *priori* prejudice is affected by the issue of whether true beliefs are better than false ones. As noted earlier, there is no requirement that this be so, but a weaker requirement is nonetheless entailed by (1) and (2):

3. If \( w_{\text{Learn}} > w_{\text{Prej}(A)} \), \( w_{\text{Prej}(\neg A)} \), then \( (b_1 + c_1) \) and \( (b_2 + c_2) \) must differ from 0 and must have the same sign.

If a proposition or its negation is of no importance \( (b_i + c_i = 0) \), then learning cannot be superior to both forms of prejudice. And if \( (b_1 + c_1) \) is positive and \( (b_2 + c_2) \) is negative, then it is better to believe proposition \( A \) whether or not \( A \) is true. When the pay-
offs are structured in this way, selection will favor some form of a priori prejudice over learning.

Proposition (3) describes two circumstances in which learning may be superior to a priori prejudice. In the first, true belief is always more advantageous than false belief. In the other, false belief is always better than truth. But propositions that don’t matter, and ones that you are better off believing regardless of whether they are true, are ruled out.

Strictly as a matter of convenience, I will assume in (most of) what follows that true belief is better than false belief. This assumption is not substantive, since counterparts of the results I will derive could be obtained under the opposite assumption.

Given the assumption that \( (b_1 + c_i) \) and \( (b_2 + c_i) \) are both positive, we can obtain from (1) and (2) a necessary and sufficient condition for learning to be better than both forms of a priori prejudice:

\[
\begin{align*}
4) \text{If } (b_1 + c_i) \text{ and } (b_2 + c_i) \text{ are both positive, then } w_{\text{learn}} &> w_{\text{prej},n} \text{ if and only if } \frac{p(1-a)}{(1-p)n} > \frac{(b_2 + c_i)/(b_1 + c_i)}{a p / [(1 - n)(1 - p)]}. \\
\text{From the criterion provided by proposition (4), we can extract a result concerning the values of the error probabilities } a \text{ and } n \text{ associated with the learning mechanism. Notice that it follows from (4) that} \end{align*}
\]

\[
\text{If } (b_1 + c_i) \text{ and } (b_2 + c_i) \text{ are both positive, then } w_{\text{learn}} > w_{\text{prej},n} \text{ only if } \frac{p(1-a)}{(1-p)n} > \frac{a p / [(1 - n)(1 - p)]},
\]

which simplifies to

\[
(5) \text{If } (b_1 + c_i) \text{ and } (b_2 + c_i) \text{ are both positive, then } w_{\text{learn}} > w_{\text{prej},n} \text{ only if } 1 > a + n.
\]

A necessary condition for learning to outperform prejudice is that the sum of the error probabilities associated with learning be less than unity. Learning demands at least this minimal degree of sensitivity. Proposition (5) says that learning is superior to prejudice only if believing a proposition is correlated with the proposition's being true:

\[
\text{Pr(Believe } A_i/ A) > \text{Pr(Believe } A_i/ \text{not-} A_i), \quad i = 1,2.7
\]

It is interesting that the necessary condition described by (5) is not sufficient for learning to be fitter than a priori prejudice. Even when the learning device effects a correlation between belief and truth, it still may not make sense to deploy the learning device.8

In addition to the requirements given by propositions (3) and (5), a third necessary condition must be satisfied, if learning is to be fitter than a priori prejudice:

\[
(6) \text{If } w_{\text{Learn}} > w_{\text{prej},n} \text{ and } A \text{ is more important than not-} A, \text{ then } \frac{p(1-a)}{(1-p)n} > \frac{a p / [(1 - n)(1 - p)]}.
\]

There is no advantage to deploying a learning device to reach a verdict on propositions that must be true, or that have no chance of being true.

The concept of a proposition's importance allows us to deduce which a priori prejudice is fitter:

\[
(7) \text{If } w_{\text{prej},n} > w_{\text{prej},n} \text{ and if and only if } \frac{p(b_1 + c_i)}{(1 - p)(b_2 + c_i)} > \frac{a p / [(1 - n)(1 - p)]},
\]

If one is going to choose an a priori prejudice, the better choice is the proposition with the greater expected importance. Notice that even if proposition \( A \) cannot fail to be true, it is not inevitable that it should be an a priori prejudice. There is no advantage in having as an a priori prejudice a proposition that is of no importance, even if that proposition must be true.

Combining propositions (6) and (7), we may conclude that necessary truths that have some degree of positive importance will be a priori.9 In contrast, it is not inevitable that a priori beliefs should be necessary.

We next turn to the issue of reliability. Proposition (1) entails the following:

\[
(5) \text{If } w_{\text{Learn}} > w_{\text{prej},n} \text{ and } A \text{ is more important than not-} A, \text{ then } \frac{p(1-a)}{(1-p)n} > \frac{a p / [(1 - n)(1 - p)]},
\]

Notice that

\[
(7) \text{If } w_{\text{prej},n} > w_{\text{prej},n} \text{ if and only if } \frac{p(b_1 + c_i)}{(1 - p)(b_2 + c_i)} > \frac{a p / [(1 - n)(1 - p)]},
\]

This means that

\[
(8a) \text{If } w_{\text{Learn}} > w_{\text{prej},n} \text{ and } A \text{ is more important than not-} A, \text{ then } \frac{p(b_1 + c_i)}{(1 - p)(b_2 + c_i)} > \frac{a p / [(1 - n)(1 - p)]}.
\]

This is, if learning is fitter than prejudice, then the learner's reliability with respect to the less important proposition must exceed the threshold specified. And if the two propositions are equally important, the learner must be more than minimally reliable with respect to both, if learning is to be better than prejudice.

Proposition (8a) is interesting because we began our discussion of learning by focusing on the issue of sensitivity, which was defined as a world-to-head relation. For learning to be superior to prejudice, we now find that a head-to-world relation also must obtain.

Learning and prejudice

55
From a biological point of view

This head-to-world conclusion is relevant to the subject of philosophical skepticism. Skepticism comes in many forms; the one that is pertinent here inquires why we should think that the beliefs an organism forms on the basis of sensory evidence are at all reliable. When an organism looks out into a field and thinks "there is a tiger," what is the probability that what the organism thinks is true? The Cartesian skeptic asks us to answer this question by appeal only to propositions that are absolutely certain. But in a more naturalistic vein, we might consider whether evolutionary considerations entail that the organism's belief is probably correct. The hope for a result of this sort may seem dim, given that Mother Nature usually goes in for comparative, not absolute, judgments; the criterion for a trait to evolve under natural selection is that it be fitter than the competition; normally, it need not satisfy any absolute requirement of adaptedness. This may encourage us to imagine that there is no limit to how unreliable a perceptual device can be and still evolve.

Proposition (8a) shows that this pessimism is exaggerated. Consider an organism that decides whether a tiger is present by consulting the deliveries of a learning device. It is often plausible to assume that for this strategy to have evolved, it must have done better than either form of a priori prejudice. This doesn't quite entail that the organism's positive and negative tiger beliefs are probably correct. What we find is something more modest, though it is interesting nonetheless: The less important proposition in the pair (a tiger is present, no tiger is present) is probably true each time it is believed.

We have just seen that the model under consideration says that when learning is fitter than prejudice, learners must be more than minimally reliable. In fact, proposition (8a) slightly understates the relevant results. No matter how propositions $A$ and not-$A$ compare in terms of their relative importance, some consequence about the reliability of learners is entailed. A fuller statement of the idea is

(8b) If $w_{\text{learn}} > w_{\text{prej}(A)}$ and $w_{\text{prej}(not-A)}$, then:
- $Pr(\text{not}-A | \text{Believe not}-A) > .5$ if $A$ is more important than not-$A$,
- $Pr(A | \text{Believe A}) > .5$ if $A$ is less important than not-$A$, and
- $Pr(\text{not}-A | \text{Believe not}-A) > .5$ and $Pr(A | \text{Believe A}) > .5$ if $A$ and not-$A$ are equally important.

Suppose that learning evolves to fixation (100 percent representation in the population) because it is fitter than both forms of a priori prejudice.

Learning and prejudice

Proposition (8b) says that if we look at the population after learning has gone to fixation, we will find that the organisms are reliable with respect to one or the other or both of the propositions (A and not-$A$) that they find themselves believing.

Now let us consider the symmetric question: If some form of a priori prejudice is fitter than learning, does it follow that the prejudiced belief will probably be true? Interestingly, the answer is no. Let us consider a case in which $w_{\text{prej}(A)} > w_{\text{prej}(not-A)} > w_{\text{learn}}$. Learning is least fit of all, we will suppose, because the error probabilities ($a$ and $n$) of the learning device are large. And proposition (7) specifies a criterion for the first prejudice to be fitter than the second; if $w_{\text{prej}(A)} > w_{\text{prej}(not-A)}$, then $p(b_1 + c_1) > (1 - p)(b_2 + c_2)$. The point to notice so far is that this last inequality places no lower bound on the value of $p$.

Someone with the a priori prejudice that proposition $A$ is true will always believe that proposition. For such individuals, $Pr(A | \text{true} / \text{Believe that } A) = p$. So even if believing $A$ as a matter of a priori prejudice is the fittest of the three strategies, nothing follows about the probability that this belief will be true.

To be sure, we can obtain a somewhat limited result concerning the reliability of a priori prejudice. Proposition (7) tells us straightforwardly that

(8c) If $w_{\text{prej}(A)} > w_{\text{prej}(not-A)}$, $w_{\text{learn}}$, and $A$ is not more important than not-$A$, then $p > .5$.

However, the point is that if $A$ is the more important proposition and $\text{Prej}(A)$ is the fittest of the three strategies, nothing follows about the reliability of the prejudice that evolves.

This asymmetry between (8b) and (8c) may have some relevance to the perennial antagonism between empiricism and rationalism. Is experience or a priori reflection a more reliable guide to the way the world is? The present model inclines toward the empiricist answer. If learning and prejudice compete against each other and learning evolves, then the deliveries of the learning device will be more than minimally reliable. However, if the competition is resolved by having an a priori prejudice evolve, nothing follows concerning the prejudice's reliability. This does not mean that the prejudice must be unreliable. But it does suggest that if learning is to evolve, it must pass a more stringent test than the prejudices with which it competes.

56

57
From a biological point of view

5. DISCUSSION

It is important that propositions (5) and (8b) not be overinterpreted. They describe minimum levels of sensitivity and reliability that a learning device must satisfy if it is to be fitter than the strategy of a priori prejudice. This, by itself, does not guarantee that the learning devices used by extant organisms live up to these minimum standards. To begin with, it is not automatic that these learning devices evolved in a population in which they competed with a priori prejudice. In addition, evolutionary models describe ancestral conditions, which may not be the same as the conditions that are in place today. Even if our ancestors excelled at identifying tigers, this says nothing about how well we do in recognizing Toyotas. And finally, the model shows that learning will be fitter than a priori prejudice only if the truth value of a belief makes a difference to the survival and reproduction of the believer. Although this may be intuitive for the contrast between believing that a tiger is present and believing that no tiger is present, it is less than obvious for other pairs of contrasting alternatives.

The concept of learning addressed in this model is a very limited one. The model describes learners as developing beliefs in response to environmental cues; they do not have access to the successes and failures they encountered on previous occasions. The individuals described here learn from their present environment, not from their experiences in the past. Obviously a more complete exploration of the advantages of learning would have to take these diachronic issues into account.12

There is one more limitation of this model that I should mention. The model treats each proposition as a separate problem. It predicts how a given proposition will be handled by attending to the proposition's importance and to the sensitivity of the learning device at hand. But it is questionable, especially in the case of our own species, that each proposition involves an independent problem.

Here I am alluding to the systematic quality of our cognitive apparatus. Perhaps the fitness of our ancestors was affected by their ability to tell whether a tiger is nearby. And perhaps it made no difference whether they could tell whether tiger lilies are at hand. However, it does not follow from this that the two propositions will be handled differently. For it may be true that any learning device that excels on the first task will probably excel on the second. If so, we may possess learning devices that deliver verdicts on propositions of no importance.13 This spin-off scenario for the evolution of cognitive capacities is important to consider, but it is not reflected in the present model.

Learning and prejudice

With these caveats in mind, we can summarize what the model says by asking when a priori prejudice makes sense. The following three considerations help make it better for an organism to believe proposition A a priori, rather than to decide whether A is true by learning:

(i) The error probabilities (a and n) of the learning device are large.
(ii) A is probably true (p is large).
(iii) A is more important than not-A (b₁ + c₁ > b₂ + c₂).

Symmetrically, we can summarize the factors that help make learning a better strategy than either type of a priori prejudice:

(iv) The error probabilities (a and n) of the learning device are small.
(v) A and not-A have middling probabilities (p is neither large nor small).
(vi) A and not-A are about equally important (b₁ + c₁ ≈ b₂ + c₂).

These factors are not separately necessary, but are considerations that promote the relevant fitness differences, all else being equal. Their interaction is depicted in Figure 3.1.

We can use these results to predict which propositions an organism will probably approach by learning, and which by a priori prejudice. If we assume that these strategies have been shaped by natural selection as modeled here, we can expect organisms to treat different propositions differently. We would not expect organisms to be pure tabula rasa; a priori prejudice will probably find some role in the organism's cognitive economy. Propositions that are important, hard to learn, and usually true can be expected to be furnished as a priori prejudices.

As a first approximation, the principle of empiricism can be taken to assert that sensory experience is the only source of information about the world.14 Those attracted by empiricism are often stymied by the problem of how the principle can be defended while remaining true to empiricist tenets. Can an empirical defense of the principle be given that avoids begging the question? If the present model is any guide, this hard question need not detain us. The reason is that the theory of evolution provides an empirical reason to doubt that empiricism is correct.15

6. AGNOSTICISM

In the above model, the only epistemic attitudes the organism can take to a given proposition is believing it or believing its negation. However,
Learning and prejudice

A third option is worth considering; the organism can suspend judgment and be agnostic.

The question to investigate is how this third option is related to the two already discussed. I will assume that the payoffs are structured as follows:

\[
\begin{array}{c|c|c}
A \text{ is true} & A \text{ is false} \\
\hline
\text{Believe } A & x + b_1 & y - c_2 \\
\text{Agnosticism} & x & y \\
\text{Believe not-A} & x - c_1 & y + b_2 \\
\end{array}
\]

It is an open question whether these three epistemic attitudes are behaviorally different in a given setting. Maybe organisms who believe that a tiger is nearby behave differently from those who believe that none is present, but those with the latter belief behave exactly like organisms who have no beliefs on the matter, one way or the other. This could be accommodated in the above model by setting \( c_1 = b_2 = 0 \). Notice also that agnosticism may have different consequences for the organism, depending on whether \( A \) is true or false.

We now can calculate how agnosticism compares with the two forms of a priori prejudice:

\[
\begin{align*}
(9) \quad & w_{\text{Prej}(A)} > w_{\text{Agnostic}} \text{ if and only if } pb_1 > (1 - p)c_2, \\
(10) \quad & w_{\text{Prej}(\text{not-A})} > w_{\text{Agnostic}} \text{ if and only if } (1 - p)b_2 > pc_1.
\end{align*}
\]

Unsurprisingly, a prejudice is better than agnosticism if the expected benefit of the prejudice’s being true exceeds the expected penalty of the prejudice’s being false. Notice that it is possible for both forms of a priori prejudice to be superior to agnosticism. Simply let \( b_1 > c_2 \) and \( b_2 > c_1 \). Nonetheless, if \( A \) is vastly more probable than its negation, one might expect Prej(\(A\)), but not Prej(\(\text{not-A}\)), to be better than agnosticism.

The next issue concerns the relation of learning and agnosticism. The relevant proposition is

\[
\begin{align*}
(11) \quad & w_{\text{Learn}} > w_{\text{Agnostic}} \text{ if and only if } \\
& pb_1(1 - a) + (1 - p)b_2(1 - n) > pc_a + (1 - p)c_n.
\end{align*}
\]

Just as in propositions (9) and (10), belief based on learning is superior to agnosticism precisely when the expected benefit of true belief exceeds the expected cost of false belief.\(^{16}\)

It is worth pointing out that the present treatment of agnosticism does not involve changing how the learning strategy was described earlier. Learners believe or disbelieve the proposition at issue; they do not
From a biological point of view
suspend judgment. A more sophisticated conception of learning would include this third option. However, a model of this type is beyond the scope of the present paper.

7. INNATENESS AND A PRIORICITY

What does it mean for a proposition to be a priori? A standard definition is that the proposition can be learned or justified by reason alone, once the constituent concepts are in hand. To say that bachelors are unmarried is an a priori proposition doesn’t mean that people are born knowing that this proposition is true. It may be that the concept of bachelorhood has to be acquired. In addition, a prioricity has to do with how a proposition could be learned or justified, not with how it was learned or justified in fact. For example, the nineteenth-century mathematician Plateau solved the problem of determining the surface of least area that fills a given contour by dipping a wire frame into a tub of soap suds (Courant and Robbins 1969). In some sense, Plateau could have figured this out by rational alone and so the proposition in question is a priori; yet his de facto method was empirical.

Although some philosophers maintain that all analytic truths are innate, I see no reason to follow them here. Perhaps many a priori truths are acquired by social learning. For example, maybe almost all people who believe the prime number theorem (that there is no largest prime) come to this belief by learning it from others. And maybe people who believe that all bachelors are unmarried do so because of experiences they have had in language learning.

So the real subject of this paper is innateness, not a prioricity. However, a word of clarification is needed in connection with the concept of innateness.

Some songbirds produce the song appropriate for their species only when they hear it first. Other species of songbird are able to produce their characteristic song even when reared in silence. And a third type of songbird will produce the song unique to its species only when it hears some song or other; even when it hears a quite different song from a different species, this will trigger the song of its own species (Gould and Marler 1991).

Should we apply the concept of “innateness” to this third case? Is the song learned from experience or is the song “performed” inside the bird, only awaiting an environmental trigger to emerge? I am skeptical that this preformationist alternative makes literal sense. Surely it makes no sense for many nonmental traits. For example, human beings grow pubic hair when they reach adolescence, but it would be absurd to say that pubic hair is “in” human babies from the time they are born.

In any event, the model provided in this paper does not require that behaviors be dichotomized in this way. We needn’t decide whether to classify the bird song as learned or innate. Rather, what is important is that we can say which environmental conditions a behavior or a belief depends upon. A phenotype can be more or less conditional on the state of the environment. It may emerge whether or not the bird hears anything at all. Or it may require hearing some sort of bird song. Or it may require hearing a particular sequence of notes. And so on.

The fundamental problem addressed here is whether a belief (or, more generally, a phenotype) should be conditional on some environmental cue. For the sake of convenience, I have classified beliefs that are relatively invariant over the organism’s experiences as “a priori prejudices” and ones that depend on the details of experience as “learned.” However, the underlying reality is a continuum.

8. WHEN WILL KNOWING EVOLVE?

Let us say that an organism knows that proposition A is true precisely when the organism truly believes the proposition and Pr(A is true/Believes that A) = 1.0. Knowledge, so defined, does not require the ability to formulate a proof that A is true, nor does it require that A be introspectively beyond doubt. Rather, knowledge is just perfectly reliable true belief.

It is quite clear from the preceding model that belief fixation devices, whether they involve learning or a priori prejudice, need not provide knowledge if they are to evolve. For a strategy to evolve, there is no requirement that Pr(A is true/Believes that A) = 1. Still, we can investigate what must be true for knowledge to arise as a by-product of selection. What must be true if an evolved belief fixation device is to give us knowledge?

We can divide this question in two. First, let us suppose that learning is fitter than a priori prejudice as a means to deciding whether proposition A is true. This means that learning will evolve to fixation. We now can use Bayes’ theorem to describe what knowing requires for the resulting organisms:
From a biological point of view

\[ \text{Pr}(A \text{ is true } / \text{Believes that } A) = \frac{\text{Pr}(\text{Believes that } A \text{ is true} \mid A \text{ is true}) \text{Pr}(A \text{ is true})}{\text{Pr}(\text{Believes that } A \text{ is true})} = \frac{(1 - a)p}{(1 - a)p + n(1 - p)} \]

If it follows that

(12) Learners know that \( A \) when they believe \( A \) if and only if \( n = 0 \) or \( p = 1 \).

We can ask the same question about the \textit{a priori} prejudice that \( A \) is true. Let us suppose that this \textit{a priori} prejudice is the fittest strategy and that it therefore evolves to fixation. If we examine an organism in the resulting population and ask whether its belief that \( A \) counts as knowledge, we again can use Bayes’ theorem to obtain an answer. For convenience, I will represent the properties of the prejudiced individual in terms of the error probability \( a \):

\[ \text{Pr}(A \text{ is true } / \text{Believes that } A) = (1 - a)p, \text{ if the individual believes } A \text{ as an } \textit{a priori} \text{ prejudice}. \]

If it follows that

(13) Prejudiced individuals know that \( A \) when they believe \( A \) if and only if \( a = 0 \) and \( p = 1 \).

The contrast between propositions (12) and (13) is of some interest. The criterion for empirical knowledge is a disjunction, whereas the criterion for \textit{a priori} knowledge is a conjunction. This suggests that \textit{a priori} knowledge may be “harder” to come by than empirical knowledge, though the conditions on both are quite formidable. It is also worth noting that sensitivity to proposition \( A \) (represented by the value of \( a \)) figures as a necessary condition in the criterion for \textit{a priori} knowledge, but plays no role at all in the criterion for empirical knowledge. Symmetrically, sensitivity to \( \neg A \) (represented by the value of \( n \)) is relevant to the question of whether the organism has empirical knowledge that \( A \) is true, but is irrelevant to the question of whether the organism knows \( A \) as a matter of \textit{a priori} prejudice.\(^{23}\)

Propositions (12) and (13) make no mention of costs and benefits. These quantities are obviously important in determining which strategy evolves. But once a strategy has evolved, the question of whether it produces knowledge has nothing to do with what is good for the organism.

---

**Learning and prejudice**

9. WHEN IS IT WORTH BEING MORE SENSITIVE?

The model described so far assesses the fitness of a strategy by mapping the costs and benefits of the beliefs that the strategy produces. However, the fitness of a piece of machinery – cognitive or otherwise – depends not only on its products, but on the “internal costs” of building and maintaining the machinery itself. For example, if an organism is going to learn whether proposition \( A \) is true, it must be equipped with sensors and with central processors that allow it to interpret the testimony of its senses. These can cost calories to build and maintain. By the same token, if an organism is going to believe proposition \( A \) as a matter of \textit{a priori} prejudice, the organism will require machinery that allows it to do this, and this machinery will impose an internal cost as well. One should not assume that these costs are irrelevant to the problem of comparing the adaptive advantages of different strategies.

Taking account of internal costs is especially important when one considers different learning strategies. Suppose an old learning strategy has an internal cost of \( I_{\text{old}} \) while a new strategy has an internal cost of \( I_{\text{new}} \). If \( a_{\text{old}}, n_{\text{old}}, a_{\text{new}} \), and \( n_{\text{new}} \) are the error probabilities associated with the old and new mechanisms, respectively, we can say when the new mechanism is fitter than the old:

\[ w_{\text{Learn-new}} > w_{\text{Learn-old}} \text{ if and only if} \]

\[ I_{\text{old}} - I_{\text{new}} > p(a_{\text{new}} - a_{\text{old}})(b_1 + c_1) + (1 - p)(n_{\text{new}} - n_{\text{old}})(b_2 + c_2). \]

If the new mechanism has a higher internal cost than the old one, then at least one of the error probabilities of the new device must be lower than the counterpart error probability of the old device. The exact reduction that is required depends on the expected importance of the two propositions. This result underwrites the intuitive idea that greater precision in belief formation (or in phenotypic response generally) is sometimes not worth the energetic candle.\(^{24}\) Improvements in fitness sometimes come by adding new equipment; at other times, the advantage goes to leaving costly equipment behind.

When learning and \textit{a priori} prejudice are compared, it is tempting to assume that the former must impose a higher internal cost than the latter. Whether this intuition is sustainable as a generalization about cognitive adaptation is an interesting problem. However, it is not always true when a facultative adaptation is compared with one that is obligatory. Consider an example of the so-called Baldwin effect.\(^{25}\) Animals are often born with calluses on their palms and feet. Yet it is clearly possible to form calluses solely as a facultative response to friction.
From a biological point of view

Why are calluses given to us, so to speak, as a priori prejudices, rather than as something that we have to acquire?

If the trait to consider is a healthy (uninfected) callus, then the facultative acquisition of this trait has nontrivial error probabilities; there is a serious chance that friction to the skin will fail to produce a healthy callus. What is more, in most environments, an organism's palms and feet are abraded, and in such environments, it is important that the organism have calluses. Plausibly, the internal cost of making calluses innate is higher than the internal cost of having them be acquired by way of friction to the skin after the organism is born. So the trait became obligate, rather than facultative, even though the internal cost of the a priori prejudice is higher than the internal cost of the learning device.

10. CONCLUDING REMARKS

As noted at the outset in connection with the polar bear's fur, there is nothing especially cognitive about the model described here. Polar bears can either have thick fur unconditionally, or they can vary the thickness of their fur in accordance with the deliverances of some measurement device that has specifiable error characteristics. Belief never enters into the picture. It is arguable, though it is not obvious, that the polar bear's problem is fundamentally different from that of a cognitive agent. Perhaps beliefs are products of cognitive strategies in a way that fundamentally differs from the way that fur is a product of thermal adaptation. I see no reason to believe this, but it is certainly a problem worth exploring.

In assimilating the problem of learning and a priori prejudice to the polar bear's fur, I have set to one side the general problem of the adaptive significance of cognition. I have not asked: Given that a phenotype is to be facultative (or obligate), why should it be placed under the control of a cognitive mechanism? Phenotypes are often dependent on environmental conditions without mental representations entering the story. Why belief should play a role is an interesting question, but it is not the one I have addressed. Rather, I have assumed that the phenotype formed by a strategy is a belief; my question has been whether the formation of the belief should be conditional on the state of the environment.

It will be evident that I have approached the issue of learning and innateness from an angle that is somewhat nonstandard. Psychologists have considered behavioral evidence for and against hypotheses of innateness (e.g., via poverty of the stimulus arguments). Philosophers have been attracted by issues of conceptual clarification: What does it mean to say that an agent believes some proposition innately, and how is this different from saying that the individual has the innate capacity to acquire the belief? For these epistemic and semantic questions, I have substituted a question of functional utility: When is it advantageous to believe a proposition innately and when is it better to have belief depend on a learning process? This evolutionary question does not replace the two more traditional queries, but I believe it has sufficient interest to merit investigation on its own.

NOTES

1. The quantifier “always” ranges over possible experiences. Also, I say that organisms conform to rules, rather than follow rules, because I don't want to require that an organism cognitively represent the strategy it uses. In the same sense, planets conform to, but do not follow, the laws of motion. For convenience, I will sometimes refer to the proposition that $A$ is false as $\neg A$. I owe this head/world terminology to Godfrey Smith (1991), who credits it to Field (1990). The present essay has several points of contact with Godfrey Smith's, which provides a definition of “optimal signal detection.”

2. More generally, note that Bayes' theorem entails that

$$Pr(\text{Believe } A | A) = \frac{Pr(A)Pr(\text{Believe } A)}{Pr(A)}.$$

The left-hand term describes the agent's sensitivity, while the conditional probability on the right describes the agent's reliability. This shows why high sensitivity does not entail high reliability, or conversely.

3. Just as would be the case in a model of the advantage of phenotypic plasticity, the probability $p$ should be understood objectively, not as reflecting the degree of belief, if any, of the organism.

4. If $(b_1 + c_1)$ and $(b_2 + c_2)$ are both negative, then the inequalities in the criterion would both reverse.

5. The more general point, of which (5) is an instance, is that learning is better than prejudice only if the learning device associates advantageous belief states with conditions that obtain in the world. It is a separate matter whether the organism is better off believing $A$ or not-$A$, when $A$ is true. And this for reasons quite separate from the matter of “internal costs,” to be discussed in Section 9.

6. Of course, an organism can't believe a given proposition if it lacks the concepts out of which the proposition is constructed. So it is possible for
From a biological point of view

there to be important necessary propositions that an organism fails to
believe, and so fails to have as \textit{a priori} prejudices.

10. This is the standard assumption in phenotypic models of selection; it can
be false if the genetic system "gets in the way." See Section 5.2 of Sober
(1993) for discussion of this point.

11. We find here a pale reflection of Donald Davidson's (1984) claim that
most of our beliefs must be true. Of course, the argument I have presented
is entirely different from Davidson's; nor is the conclusion as general as
his. However, it does show that, in expectation, most of the beliefs that
fall in a specific class (i.e., ones that are less important than their nega-
tions and are acquired by a learning mechanism that evolved in the man-
ner specified) are true. This issue is explored more fully in Essay 4.

12. See, for example, the learning strategies discussed in Pulliam and Dun-

13. It also is possible that different "modules" should deliver different verdicts
on the very same proposition. A proposition \( P \) may be enshrined as an \textit{a priori}
prejudice and yet a learning device may tell the organism that \( P \) is
false. Just as two environmental indicators may come into conflict, so an
environmental indicator may conflict with an \textit{a priori} prejudice.

14. I owe this informal rendition of the principle to Bas Van Fraassen. Not-
wittingly the fact that the concept of "information" needs to be clari-
fied, this statement of the principle is a good place to begin.

15. In spite of this negative verdict, I believe that there are important empiri-
cist ideas that ought to be preserved in an adequate epistemology. This
point is elaborated in Essay 6.

16. In "The Will to Believe," William James (1897) remarks that "he who says
'Better go without belief forever than believe a lie' merely shows his own
preponderant private horror of becoming a dupe." In light of the above
model, it is no surprise that Mother Nature did not equip us with the
Cartesian policy of remaining agnostic toward all propositions except
those that are subjectively (or objectively) certain.

17. Indeed, it may be better still to move to a degree of belief conception, in
which there are far more than three epistemic attitudes to consider. In
doing so, the problem of how to assign a degree of belief to a proposition
will be separated from the problem of how to behave, since action will be
a joint function of probabilities and utilities (and a decision rule). The
present model, in which the agent either believes or disbelieves a proposi-
tion, posits too close a connection between belief and action.

A further advantage of the degree of belief approach is that the model
of phenotypic plasticity described by D. W. Stephens (1989) can be applied
to the case of belief. Stephens' approach, which is based on the work
of the economist J. P. Gould (1974), requires that the phenotypes consid-
ered be located on a metric. Although it is difficult to see how to do this
for a set of mutually exclusive beliefs, there is no such problem when one

Learning and prejudice

considers different degrees of belief that an agent might have with respect
to a single proposition.

It is worth mentioning that the Stephens-Gould definition of the "value
of information" entails that learning is never worse than \textit{a priori} prejudice.
Of course, no such result follows from the framework developed here.
This clash is only apparent; it arises from the way that Stephens and
Gould define the concept of a learning strategy.

18. This is the position taken by Fodor (1975). At most, Fodor's argument
shows that concepts cannot be \textit{learned}. But that isn't enough to show that
they are innate. When I go to Florida, I acquire a suntan, but not by
learning.

19. Fodor (1981) argues that the difference between learning and triggering
is fundamental to the problem of innateness.

20. If talk of innatism sometimes carries with it an indefensible commitment
to preformationism, scientists avoid this vestige of a bygone age by talking
about \textit{norms of reaction} and about the issue of heritability. I provide an
elementary explanation of these ideas in Sober (1993).

21. This is the essence of various "externalist" theories of knowledge. See, for
example, Dretske (1981).

22. This, I take it, is the main conclusion of Godfrey Smith's (1991) discussion
of signal detection theory.

23. Of course, the fact that different conditions pertain to \textit{a priori} and to em-
pirical knowledge in no way depends on the evolutionary considerations
explored here. The difference is a simple consequence of applying Bayes' the-
orem to the reliabilist definition of knowledge.

24. Stich (1990) argues that natural selection cannot be expected to reduce
the error probabilities of a perceptual device to zero. The present model
reinforces Stich's point. Even if true belief is a component of fitness, it is
not the whole story.

25. This is discussed by C. D. Waddington (1957) and by G. C. Williams
(1966). I examine Williams' analysis of this phenomenon in Sober (1966),
Section 6.2.

26. Indeed, the model developed in Moran (1992) concerning the evolution
of polyphenism is essentially the same as the one presented here. However,
since Moran's interests are not in cognitive strategies \textit{per se}, the inter-
pretive points she makes do not overlap much with the ones I have developed.

27. I use the term "measurement device" in roughly the way a physicist
would. Any state of the polar bear's body that is \textit{correlated} with the ambient
temperature is a measurement device in the sense required.