

## Précis of *Evidence and Evolution: The Logic behind the Science*

ELLIOTT SOBER

*University of Wisconsin, Madison*

*Evidence and Evolution* has four chapters: (1) Evidence, (2) Intelligent Design, (3) Natural Selection, and (4) Common Ancestry. The first chapter develops tools that are used in the rest of the book, though more ideas about evidence are added along the way.

The first chapter gives a brief introduction to Bayesianism and Likelihoodism. Bayesianism is based on using Bayes' Theorem, either to compute the probability that hypotheses have in the light of given evidence, or, more modestly, to say which of the competing hypotheses is most probable. The central concept of Likelihoodism is the Law of Likelihood:

(LoL) Evidence  $E$  favors hypothesis  $H_1$  over hypothesis  $H_2$  if and only if  $\Pr(E|H_1) > \Pr(E|H_2)$ .

Note that the right-hand side of the (LoL) describes the probability of the evidence according to the different competing hypotheses (the symbol “|” means *given*), not how probable the hypotheses are, given that evidence. The (LoL) isn't a theorem; rather, it is a proposed explication. It is something I repeatedly put to work in *Evidence and Evolution*—comparing evolutionary theory and creationism (Chapter 2), comparing natural selection and drift (Chapter 3), and comparing common and separate ancestry (Chapter 4).

After the introduction to Bayesianism and Likelihoodism, Chapter 1 continues with a discussion of Frequentist approaches to theory evaluation. I am critical of two prominent Frequentist tools—significance tests and Neyman-Pearson hypothesis testing. I locate the former procedure under the heading of *probabilistic modus tollens*; this is the idea that we should reject a hypothesis because it says that what we observe

has a very low probability. I then give a sympathetic presentation of another Frequentist idea, the Akaike Information Criterion (Akaike 1973); AIC forms part of the statistical subject of model selection theory.

In Chapter 1, I endorse a pluralistic outlook—Bayesianism is fine in some inference problems, likelihoodism in others, and AIC in still others. These problems differ from each other both in terms of the kind of information one has available and in terms of one's goals. With respect to information: Bayesianism is a legitimate framework when the values assigned to prior probabilities and to likelihoods have an objective justification, likelihoodism comes into its own when priors cannot be justified in this way though likelihoods can, and AIC makes sense when neither priors nor likelihoods can be justified. With respect to goals: the goal of Bayesian inference is to determine how probable this or that hypothesis is (or, more modestly, which of the competing hypotheses is most probable), the goal of likelihood inference is to determine which of the competing hypotheses is favored by the evidence at hand, and the goal of AIC is to determine which model will be most predictively accurate (where it turns out that a model known to be false can sometimes be expected to be more predictively accurate than a model known to be true). In this first chapter, I discuss some probability puzzles concerning stopping rules and how one should reason about coincidences. I also discuss how Bayesianism, Likelihoodism, and AIC are related to larger questions in the philosophy of science concerning realism, instrumentalism, and idealization. And parsimony makes its first appearance in this chapter, in a discussion of how AIC uses the number of adjustable parameters that a model deploys to help estimate how predictively accurate the model will be.

In Chapter Two, on intelligent design, I try to develop the strongest possible formulation of the design argument for the existence of God, after which I criticize the argument so conceived. Paley's (1802) analogy between a watch and the vertebrate eye is where I start. I end up interpreting the argument, not as a Bayesian argument that an intelligent designer probably made the eye, nor as an analogy argument, but as a likelihood inference. I then criticize the argument by discussing how Duhem's Thesis applies in nondeductive contexts. Duhem (1914) and Quine (1953) say that theories typically don't have observational consequences all by themselves, but need to be supplemented with auxiliary assumptions in order to do so. I argue that theories often don't confer probabilities on observations all by themselves, but need to be supplemented with auxiliary assumptions in order to do so. I then discuss what sorts of auxiliary assumptions we are entitled to add to the theories we wish to test. I argue that we can't simply invent

assumptions that lead theories to do well or poorly in the likelihood competition. Auxiliary assumptions must be justified. This allows me to criticize the likelihood version of the design argument. When Paley claims that the complex functionality of the camera eye favors the hypothesis of intelligent design over the hypothesis of chance, he is making assumptions about the putative designer's goals and abilities that he cannot independently justify. This criticism of the design argument, it turns out, has implications concerning a standard argument *against* intelligent design that biologists have frequently made. Greg Radick (2005) calls it the no-designer-worth-his-salt argument. Darwin (1859), Stephen Jay Gould (1980), and many other biologists have argued that imperfect adaptations are evidence against intelligent design. Although this is a good argument against an omnipotent God who wants above all to give organisms perfect adaptations, it does not touch other versions of ID. For example, when Gould argues that the panda's thumb is evidence against the design hypothesis, he is making assumptions about the putative designer's goals and abilities, assumptions that he has no right to make. Gould assumes that the putative designer would have had the goal of giving pandas a device for stripping bamboo that would have been far more efficient than the thumb that pandas currently have, and that the designer would have had the ability to furnish pandas with this better device. Gould falls into the same trap that tripped up Paley.

This second chapter also attempts to resuscitate the concept of testability. Philosophers nowadays often think that the search for a criterion of testability is a fool's errand, something that the overthrow of logical positivism has taught us is a pseudo-problem. I disagree. I think philosophers of science *should* try to understand what testability is and I take a shot at doing so. The chapter also considers two versions of the design argument additional to the likelihood version I just mentioned. One uses model selection ideas; the other addresses an argument that Cleanthes makes in Hume's (1779) *Dialogues Concerning Natural Religion*. The chapter also has some discussion of Michael Behe's (1996) contention that features of organisms that exhibit "irreducible complexity" cannot be explained by evolutionary processes and must be the work of an intelligent designer.

Chapter 3, on natural selection, starts with polar bears. How should we go about testing two hypotheses about why the average fur length of present day polar bears is 10 centimeters? The hypotheses I consider are Selection-plus-Drift (SPD) and Pure Drift (PD). I put this problem in a likelihood framework. We need to figure out which of these hypotheses confers the larger probability on the observed fur length. This requires auxiliary assumptions. The standards that apply to

evaluating hypotheses of intelligent design apply here as well. We can't simply invent auxiliary assumptions; rather, they must be justified. After discussing this problem and noting how hard it is to solve, I shift to another problem that I think is more tractable. Instead of trying to explain why polar bears have fur that is 10 centimeters long, we can take up the question of why bears in cold climates tend to have longer fur than bears in warm climates. I argue that it is correlations across species, rather than the trait values of a single species, that evolutionary biologists typically should aim at explaining. This leads to considerations from model selection theory and reflection on how a theory that unifies different observations can be better than a theory that disunifies them. Reichenbach's (1956) principle of the common cause then comes in for criticism. And there is discussion of how biologists use molecular data to test selection against drift and also of how they use phylogenetic information to test adaptive hypotheses. This chapter also evaluates the use that biologists make of a concept called *cladistic parsimony* to infer the character states of ancestors. This kind of parsimony differs from the one that is used in model selection theory; I look at cladistic parsimony through the lens of likelihood and also from a Bayesian point of view.

Chapter 4 is on common ancestry. I start by discussing what it means to say that all life on earth has a common ancestor. This involves the concept of *tracing-back*; it does not mean that life originated on earth exactly once, nor does it mean that only one "original progenitor" has descendants that exist now. I then turn to evidential questions. Why should the similarity of two species be evidence that they have a common ancestor? Which sorts of similarity provide stronger evidence for common ancestry and which provide weaker evidence or none at all? I make use of Reichenbach's ideas about how a common cause is probabilistically related to its separate effects (though not by endorsing his principle of the common cause) and give a likelihood treatment of these questions. Markov models of evolution are considered. The chapter concludes with a discussion of how biologists use cladistic parsimony to infer phylogenetic trees and how this parsimony procedure compares with explicitly statistical approaches that use maximum likelihood instead. Model selection ideas come in handy here.

### References

- Akaike, H. (1973): "Information Theory as an Extension of the Maximum Likelihood Principle." In B. Petrov and F. Csaki (eds.), *Second International Symposium on Information Theory*. Budapest: Akademiai Kiado, pp. 267–281.

- Behe, M. (1996): *Darwin's Black Box*. New York: Free Press.
- Darwin, C. (1859): *The Origin of Species*. London: Murray; Cambridge: Harvard University Press, 1964.
- Duhem, P. (1914): *The Aim and Structure of Physical Theory*. Princeton: Princeton University Press, 1954.
- Gould, S. (1980): *The Panda's Thumb*. New York: Norton.
- Hume, D. (1779): *Dialogues Concerning Natural Religion*. London: Penguin, 1990.
- Paley, W. (1802): *Natural Theology, or, Evidences of the Existence and Attributes of the Deity, Collected from the Appearances of Nature*. London: Rivington.
- Quine, W. (1953): "Two Dogmas of Empiricism." In *From a Logical Point of View*. Cambridge: Harvard University Press, pp. 20–46.
- Radick, G. (2005): "Deviance, Darwinian-Style—A Review of A. Lustig, R. Richards, and M. Ruse's *Darwinian Heresies* (Cambridge: CUP, 2004)." *Metascience* 14: 453–457.
- Reichenbach, H. (1956): *The Direction of Time*. Berkeley: University of California Press.

# Responses to Fitelson, Sansom, and Sarkar

ELLIOTT SOBER

*University of Wisconsin, Madison*

My thanks to Branden Fitelson, Roger Sansom, and Sahotra Sarkar for their thought-provoking comments on *Evidence and Evolution*. I respond in what follows to each of them separately.

## 1. Responses to Sahotra Sarkar

### 1.1. *Is Confirmation Contrastive?*

Sarkar says that I hold that “theory confirmation is always contrastive.” This means that if an observation confirms hypothesis  $H_1$ , the observation does so only because  $O$  and  $H_1$  are related in a certain way to a second hypothesis, call it  $H_2$ . A similar thesis pertains to testing. To test a hypothesis  $H_1$  is to test it against some alternative hypothesis  $H_2$ .

I do not endorse the contrastive thesis when  $H_1$  and  $O$  are deductively related. There is nothing wrong with *modus tollens*. If  $H_1$  entails  $O$  and  $O$  is false, this entails that  $H_1$  is false; there is no need to bring in an alternative hypothesis  $H_2$  to see this. Nor do I have a beef against *modus ponens*. If  $O$  entails  $H_1$  and  $O$  is true, you don’t need to consider an alternative hypothesis  $H_2$  to see that  $H_1$  is true. Contrastivism needs to be restricted to cases in which hypotheses and observations are not deductively connected.

Likelihoodism wears its contrastive character on its sleeve. The Law of Likelihood says

$$(\text{LoL}) O \text{ favors } H_1 \text{ over } H_2 \text{ iff } \Pr(O|H_1) > \Pr(O|H_2).$$

And likelihoodism in its usual formulations includes the insistence that the absolute value of  $\Pr(O|H_1)$  doesn’t tell you anything about whether  $O$  is evidence for or against  $H_1$  (Royall 1998, p. 67). It is the

relationship between likelihoods, not single likelihoods taken all by themselves, that are epistemically significant.

Bayesianism is contrastive too, though this is slightly less obvious. Bayes' theorem says that

$$\Pr(H|O) = \frac{\Pr(O|H)\Pr(H)}{\Pr(O)}$$

and the only hypothesis mentioned in this equation is H. But dig a little deeper and the contrastive element becomes visible. The probability of the observation,  $\Pr(O)$ , is, in fact, the average likelihood of H and not-H:

$$\Pr(O) = \Pr(O|H)\Pr(H) + \Pr(O|\text{not-H})\Pr(\text{not-H}).$$

Similarly, the Bayesian theory of confirmation

$$O \text{ confirms } H \text{ iff } \Pr(H|O) > \Pr(H)$$

is equivalent to

$$O \text{ confirms } H \text{ iff } \Pr(O|H) > \Pr(O|\text{not-H}).$$

Bayesian confirmation pits a hypothesis against its own negation (Sober 2008, p. 34). The Law of Likelihood is less restrictive. Compare the likelihoods of H and not-H if you want (assuming that both of these make sense), but you also can contrast  $H_1$  with  $H_2$  where  $H_2$  isn't the negation of  $H_1$ . Likelihoodists typically have views about the kinds of hypotheses that can figure in legitimate likelihood comparisons; I'll discuss this later.

Are there noncontrastive epistemologies for hypotheses that are related only probabilistically to the observations considered? Significance tests are an example. They are instances of what I call *probabilistic modus tollens*:

Data

$\Pr(\text{Data}|H)$  is low

H should be rejected (or: the data are evidence against H)

I criticize significance tests in *Evidence and Evolution* (pp. 49–53). All the other evidential frameworks that I canvas there are contrastive—not just Likelihoodism and Bayesianism, but the Neyman-Pearson theory

of hypothesis testing, and model selection criteria like AIC. And the multi-criteria framework for evaluating theories that Sarkar favors is contrastive as well. More on this in the next section.

### 1.2. *Is Evolutionary Theory Explanatorily Better than the Hypothesis of Intelligent Design?*

In the chapter of *Evidence and Evolution* on intelligent design, I focus mainly, but not exclusively, on a likelihood formulation of the design argument. According to this interpretation, Paley and other proponents of this argument are attempting to show that

$$\Pr(\text{organism } o \text{ has features } F \mid \text{an intelligent designer made } o) > \Pr(\text{organism } o \text{ has features } F \mid \text{mindless process } X \text{ made } o).$$

Before Darwin, the mindless process that people usually had in mind was “chance”—the Epicurean idea of particles in random motion that occasionally stick together and form stable configurations. Afterwards,  $X$  was usually the process of evolution by natural selection. The features  $F$  that usually get plugged into this schema are ones that characterize complex adaptive structures (like the vertebrate eye). My complaint about this argument is that there is no saying whether the left-hand likelihood is one or zero or something in between until specifics are provided about the goals and abilities that the putative intelligent designer would have if this being existed. One can’t just stipulate what these are; they must be based on a plausible argument that shows that if an intelligent designer made this organism, then the designer probably would have had *these* goals and abilities rather than *those*. I do not demand that we have total certainty about exactly what those goals and abilities are; what is needed is that we know enough about this to allow the likelihood evaluation to go forward (Sober 1998, pp. 142–143).

Sarkar agrees with this criticism of the design argument, but he thinks it shortchanges Darwin and Wallace, depriving them of the credit they are due. He thinks that my analysis of the design argument doesn’t show why evolutionary theory (ET) “better explains” the features of organisms than the intelligent design hypothesis (ID) does. I’m not sure what Sarkar means when he says that ET is explanatorily better than ID. He says at one point that ID’s being “intellectually incoherent” would not show that ET is explanatorily better than ID. I use the term “better explanation” in a more liberal way. If ID is incoherent and ET is not, then ET is better in a way that pertains to its ability to explain. And if ID is untestable whereas ET is not, that too is a point in favor of ET as an explanation of what we observe.

Sarkar is right that my position on Darwin's relation to the design argument has changed. I once thought that the ID hypothesis has a higher likelihood than Epicureanism, but that ET then came along and ET turned out to have a higher likelihood than ID (Sober 1990). I now think that this view of what is going on is mistaken. ID never was much of a theory—not because it makes false predictions, but because it doesn't predict much of anything at all. This defect in ID can be recognized without taking Darwin's theory into account. In spite of this change in view, I continue to subscribe to my statement in *Philosophy of Biology* that Darwin "altered the dialectical landscape" by developing his new theory (Sober 1993). Sarkar thinks my critique of the likelihood version of the designer argument in *Evidence and Evolution* requires a retraction of that comment about Darwin; I don't see that it does.

When Sarkar brings his multi-criteria framework for evaluating competing theories to bear on the three-way competition between ID, ET, and the Epicurean hypothesis that life evolves by a purely random process, he reaches the conclusion that ET has a higher likelihood than Epicureanism and, as I said, he agrees with me that the likelihood of ID can't be evaluated. He concludes that "this is what Darwin and Wallace achieved." I agree with Sarkar that Darwin constructed a theory that has a higher likelihood than Epicureanism. This *was* an achievement. But how does that show how Darwin's accomplishments bear on the credentials of intelligent design? I don't see that Sarkar has an answer to this question that differs from my own.

Sarkar describes two formats that his multi-criteria framework for evaluating theories might follow. I won't discuss the second set-up, which involves second-order probabilities. The first format involves discarding theories that can't be evaluated for their likelihoods or their prior probabilities, and then seeing if any of the remaining theories are non-dominated. When there is a three-way competition between intelligent design, purely random processes, and evolution by natural selection, we drop the first of these (because its likelihood cannot be evaluated) and then compare the second and third, because both their likelihoods and their priors can be computed. The way this approach handles ID resembles what I say in *Evidence and Evolution*. If my approach fails to describe what Darwin and Wallace achieved (because it fails to show how their theory of evolution is explanatorily better than ID), then I think the same is true of Sarkar's. However, there is a difference between Sarkar's proposal and my likelihood treatment. Sarkar wants to compare the prior probabilities of random processes and natural selection. He doesn't

say how this might be done. I am skeptical that one can provide an objective justification for ordering the priors in one way rather than another. This isn't because I think there is something special about these two theories. I think this is the typical situation for "big" theories—relativity and quantum mechanics, for example.

## 2. Responses to Roger Sansom

### 2.1. *On Rescuing the Design Argument*

Sansom discusses two versions of the Design Argument, each construed as a likelihood argument. The first he terms "non-stipulative." It has separate premises that make claims about the goals and abilities the putative designer would have if he existed. Sansom makes no objection to my criticism of this argument, but says that there is another rendition of the design argument against which my criticism has no force. This is what he calls "the stipulative theory of intelligent design". To illustrate the issues here, consider the observation that vertebrates have eyes and the following two hypotheses:

H<sub>1</sub>: An omnipotent intelligent designer wanted above all that vertebrates have eyes.

H<sub>2</sub>: Life evolved.

The likelihoods of the two hypotheses, relative to the observation that vertebrates have eyes, are:

$$\Pr(\text{vertebrates have eyes} \mid H_1) = 1.$$

$$\Pr(\text{vertebrates have eyes} \mid H_2) < 1.$$

There no longer is any need to consider auxiliary assumptions to figure out what the likelihood is of the ID hypothesis; the likelihood of H<sub>1</sub> unproblematically has a value of one.

I discuss this suggestion in *Evidence and Evolution* (pp. 131–136) and explain why I think it is a dead end. The stipulative formulation of the design argument provides a hollow victory for ID for the simple reason that two can play the same game. If we pack the observations into the ID hypothesis, why not do the same for the evolutionary hypothesis? Instead of H<sub>2</sub>, consider the following hypothesis:

H<sub>3</sub>: life evolved with the consequence that vertebrates have eyes.

If you pack in for both, both hypotheses have a likelihood of unity, which, according to the Law of Likelihood, means that the observations do not discriminate between the two hypotheses. You then may be tempted to evaluate the theories in terms of their prior probabilities, but now two problems arise. First, this shift to priors means that you are abandoning the idea that the observed features of organisms actually do some work in testing intelligent design against a competing hypothesis. Since the design argument is supposed to rest on observations, this shift to priors does not do justice to what the argument aspires to do. The second problem is that it is often impossible to provide objective justifications for assignments of prior probabilities; I think this is the case for  $H_1$  and  $H_3$ .

There is nothing special about evolution versus intelligent design here. For any observation  $O$ , a deterministic theory  $D$  that says that  $O$  had to happen will have a higher likelihood than a stochastic theory  $S$  that says that  $O$  happened with some probability less than unity. How, then, could evidence ever favor the stochastic theory over a deterministic theory? I see two solutions. The first is to break the hypotheses into pieces. This is what I did in my likelihood treatment of the design argument in which auxiliary hypotheses are separated from the hypotheses under test, and the auxiliary assumptions are required to be independently justified. This is one way to avoid trivializing the question of how observations can discriminate between competing hypotheses. Epistemological holism has been much in vogue in philosophy, thanks in large measure to the influence of Quine (1953); I think this holism is a mistake (Sober 2004). Breaking a theory into pieces and testing the pieces separately is a strategy that is widely used in the sciences. The second strategy for avoiding trivialization moves beyond the Law of Likelihood and views matters from the vantage point of model selection theory. Deterministic theories sometimes involve models that are very complex, while stochastic models are simpler. Here a model's complexity is measured by the number of adjustable parameters. This difference in parsimony is epistemically relevant, according to statistical ideas from model selection theory. I discuss this way of dealing with the competition between  $D$  and  $S$  in *Evidence and Evolution* (pp. 104–107 and pp. 177–184).

## 2.2. On Testing Pure Drift against Selection-Plus-Drift

In Chapter 3 of *Evidence and Evolution*, I discuss how one might go about testing the hypothesis of selection-plus-drift (SPD) against a hypothesis of pure drift (PD). The example I initially use to discuss this

involves the observation that polar bears now have an average fur length of (say) 10 centimeters. SPD says that there was variation in fitness in the lineage leading to present day polar bear, but population size was finite; the SPD hypothesis postulates an optimal fur length for polar bears and says that the population average evolved in the direction of that value, not with certainty, but in probabilistic expectation. PD says that there was no variation in fitness in the finite population, so trait values evolved by random walk. The question is whether the SPD hypothesis makes the observed fur length more probable than the PD hypothesis does. If this is the case, the Law of Likelihood will conclude that the observed fur length *favors* SPD over PD.

I assumed in my analysis of this problem that polar bear fur length must be between 0 and 100 centimeters. I then argued that when enough time has passed in the lineage, the SPD hypothesis predicts that the observed fur length should be “close” to the optimal fur length it postulates, whereas the PD hypothesis says that all possible fur lengths are roughly equiprobable. The SPD hypothesis will have the higher likelihood if the observed fur length is “close enough” to the optimum it postulates. I then describe various empirical facts that need to be in place to determine what “close enough” means. The main point of this exercise is to suggest that this way of thinking about testing SPD against PD is a dead end (Sober 2008, pp. 219–226). I don’t think there is much hope for fleshing out the empirical details that are needed to make the likelihood comparison work. I suggest instead that we evaluate PD versus SPD by looking at multiple species that stem from a common ancestor and at the pattern of correlation exhibited among them. Don’t ask why polar bears have a fur length of 10 centimeters. Ask instead why bears in colder climates tend to have longer fur than bears in warmer climates. This observed negative correlation between fur length and temperature has a higher probability of obtaining under the SPD hypothesis than it does under PD. Happily, you need to know far less about biological details to make this likelihood argument work.

Sansom thinks that the SPD and PD models assume that infinite time has passed. I disagree. The models describe how the probability distributions *evolve* as time marches on. Figure 1 in Sansom’s article depicts what is true after infinite time has passed, and so it provides a good approximation when there has been a lot of time between the process’s start and finish. When there is less time between start and finish, the SPD and PD hypotheses still can confer different probabilities on the observed polar bear trait value; see the figure on page 199 of *Evidence and Evolution*.

Sansom suggests that the analysis I give depends on how broad a range of possible trait values one chooses, noting that I mention the arbitrariness of the 100 centimeter cut-off in *Evidence and Evolution*

(p. 224). I assumed that 0 to 100 centimeters is the range of possible values. Will it affect the likelihood analysis of PD versus SPD if you assume, instead, that the range is 0 to 500? Sansom thinks so and that this is a problem for the conclusions I want to draw. I'm still not sure how much this shift will affect the likelihood assessment, but if it does have a significant impact, that result would bolster my contention that this format for testing PD against SPD is a dead end. A similar point applies to Sansom's comment that the PD distribution (after infinite time has passed) need not be flat. I should have pointed out in the book that whether this is so depends on the magnitude of the mutational input; with zero or near-zero mutation, the distribution will be U-shaped, with a small mutational input it will be flat, but with a large input it will be single-peaked (Durrett 2002, pp. 14–23). But here again, the point of importance is a general one: to get the SPD and the PD hypotheses to confer different probabilities on the observations, a good deal of biological information is needed.

Sansom then turns to what he takes to be a second model in *Evidence and Evolution*. In fact, the shift he is describing isn't to new models of PD and SPD, but to a *different data set* that is used to evaluate the two old models. For both problems, SPD and PD are understood in terms of the Ornstein-Uhlenbeck process that I describe. The shift is from trying to explain why polar bears have an average fur length of 10 centimeters to trying to explain why bears in colder climates tend to have longer fur than bears in warmer climates. That is, what we are trying to explain in this second problem is why the best-fitting regression line drawn through data on different bear species ( $x$  = ambient temperature,  $y$  = average fur length) has a negative slope. PD says that descendant lineages did a random walk as they diverged from their most recent common ancestor. SPD says that they, in expectation, moved in the direction of an optimality line that has a negative slope. I say in the book that SPD predicts our observation that bears in colder climates tend to have longer fur than bears in warmer climates, whereas PD predicts that the slope of the regression line through the observations has a slope of zero. Sansom agrees with this assessment of what PD predicts, but he says "it appears inconsistent with Sober's second model. A horizontal line of best fit is the expectation for an infinite set of species, but for a finite number PD predicts a normal distribution of lines of best fit that center on horizontal." My use of the word "predicts" here was a bit sloppy. The more careful statement is that SPD confers a higher probability on the observation that the slope is negative than the PD hypothesis does. If "H predicts O" entails that  $\Pr(O|H) > 0.5$ , then "prediction" is a distraction from what is crucial for likelihood comparisons (Sober 2008, p. 294).

Sansom notes that it is possible for bear species to exhibit a negative regression line if the lineages evolve away from the optima postulated by the SPD hypothesis. This is true, but it doesn't affect the likelihood comparison I am discussing—the negative slope that summarizes the observations is still more probable under SPD than it is under PD. Sansom also says that “one should not conclude that even a relatively strong correlation must be due to this particular SPD hypothesis, because there are many other potential selection pressures and constraints that may differ across species that could also result in a negative correlation.” This point is correct; it does not conflict with what I say in *Evidence and Evolution*. I there use the Law of Likelihood to evaluate two hypotheses in the light of a set of observations; the likelihood argument draws no conclusion about which hypothesis one should accept (see the separation of three questions on pp. 3–8) and the analysis does not involve surveying other possible hypotheses.

My extended discussion of fur length in bears, where one shifts from the task of explaining a single trait in a single species to a pattern in a multi-species data set, is supposed to have a general lesson. Informal discussions of evolution, both by nonbiologists and by biologists also, are often about the trait value found in a single species or higher taxonomic group. Why, for example, do vertebrates have camera eyes? If we set story-telling aside, we should look instead at the distribution of eye designs, including the state of having no eyes at all, across a more inclusive set of taxa. A time-honored problem in philosophy of science involves observing a single black raven and asking what this observation says about the hypothesis that all ravens are black. This is the wrong model for testing drift against selection.

### 3. Responses to Branden Fitelson

#### 3.1. *The Law of Likelihood is Dead. Long Live the Law of Likelihood!*

I agree with Fitelson that the Law of Likelihood has counterexamples when no restrictions are placed on the hypotheses and observations to which it applies. Here's one counterexample that I put in *Evidence and Evolution* (pp. 36–37); it was inspired by Fitelson (2007). You are Madison's top meteorologist. You use data on the current weather and a good model of how weather systems change. Based on that information, you see that it probably will rain tomorrow. Your theory plus data “favors” rain over no rain, but this isn't because rain has the higher likelihood. What favoring means here is just that

$\Pr(\text{rain tomorrow} \mid \text{data about now \& model}) > \Pr(\text{no rain tomorrow} \mid \text{data about now \& model}).$

What is not at issue here is the claim that

$\Pr(\text{data about now} \mid \text{rain tomorrow}) > (\text{data about now} \mid \text{no rain tomorrow}).$

Sometimes hypotheses are “favored” because of their probabilities, not because of their likelihoods. What I conclude from this and other examples is that the Law of Likelihood isn’t a good explication of what the word “favoring” means in English. However, this doesn’t lead me to think that the Law of Likelihood is a defective epistemological principle.

Likelihoodism is not Bayesianism. It is a philosophy that comes into its own when the hypotheses we are considering don’t have objective prior or posterior probabilities. When Eddington was using his eclipse data to test relativity theory against Newtonian theory, he was able to consider how each theory probabilifies his observations. He was not able to say which theory was rendered more probable by those observations. As this example also illustrates, likelihoodism does not aim to address cases in which the evidence deductively entails one competing hypothesis, but not the other.

Likelihoodism can be viewed as a part of Bayesianism because of the role the likelihood ratio plays in the odds formulation of Bayes’ Theorem:

$$(O) \quad \frac{\Pr(H1|O)}{\Pr(H2|O)} = \frac{\Pr(O|H1)}{\Pr(O|H2)} \times \frac{\Pr(H1)}{\Pr(H2)}$$

What this theorem says to a Bayesian is that the only way the ratio of posterior probabilities can differ from the ratio of priors is by the likelihoods differing. The more the likelihood ratio deviates from unity, the more one’s ratio of credences is transformed by the observations. For Bayesians, likelihoods provide the sole pathway by which relative degrees of certainty can be altered by observations. What the Law of Likelihood does is draw a circle around the likelihood ratio and say: *this* is how the observations are relevant to the competition between the two hypotheses.

The word “favoring” is ill-chosen as a device for pointing out the evidential significance of likelihoods. So, for that matter, is the word “discrimination.” Indeed, the whole subject was conceived in semantic sin, since Fisher couldn’t have chosen a more misleading word for

likelihood than the word he chose—namely, the word “likelihood.” However, the fact that “likely” and “probably” are synonyms in English does not mean that Fisher went wrong in arguing for the epistemological significance of likelihoods. Similarly, the role of the likelihood ratio in (O) should lead Bayesians to *embrace* the Law of Likelihood, once the bad terminology is duly recognized.

3.2. *Commutativity, the Ratio Measure of DOC, the Bridge Principle, and the Popperian Principle*

Fitelson points out that the LoL is implied by the ratio measure of degree confirmation

$$\text{DOC}_R(\text{O},\text{H}) = \frac{\text{Pr}(\text{H}|\text{O})}{\text{Pr}(\text{H})}$$

He then notes that the ratio measure is flawed. One flaw is that it entails that degree of confirmation is commutative, since

$$\text{DOC}_R(\text{O},\text{H}) = \text{DOC}_R(\text{H},\text{O}).$$

So far so good. How does this affect the Law of Likelihood? Not at all. Rejecting the ratio measure does not require you to reject everything that it entails. In fact, likelihoodism is philosophically *opposed* to commutativity and, for that matter, to computing the ratio of posterior probabilities to priors. One of the main points of that philosophy is to talk about cases in which  $\text{Pr}(\text{O}|\text{H}_1)$  and  $\text{Pr}(\text{O}|\text{H}_2)$  make sense while  $\text{Pr}(\text{H}_1|\text{O})$  and  $\text{Pr}(\text{H}_2|\text{O})$  do not. The hypotheses of interest confer objective probabilities on observations, but the observations do not confer objective probabilities on hypotheses.

Related remarks apply to Fitelson’s bridge principle:

(BP) E favors  $\text{H}_1$  over  $\text{H}_2$  iff E supports  $\text{H}_1$  more than E supports  $\text{H}_2$ .

Fitelson says that this principle “facilitates a comparison of Likelihoodist versus Bayesian explications of the favoring relation.” It does—if you are a Bayesian. As Fitelson says, “from a Bayesian point of view, the debate about (LoL) is really just a debate about the proper measure of degree of confirmation.” This is true, but likelihoodists do not view the (LoL) in this way. Likelihoodists typically don’t think that the favoring relation involves a comparison of the degree to which E supports  $\text{H}_1$  and the degree to which E supports  $\text{H}_2$ . Favoring, for them, is an irreducibly three-place relation; it doesn’t decompose into a

two-place relation between  $H_1$  and  $E$  and a two-place relation between  $H_2$  and  $E$ .

Fitelson describes a Popperian Principle about favoring and says that the Law of Likelihood (LoL) is meant to be a probabilistic generalization of (PP). Here's the principle:

(PP) If  $H_2$  entails not- $E$  but  $H_1$  does not entail not- $E$ , then  $E$  favors  $H_1$  over  $H_2$ .

Fitelson endorses this principle, saying that it expresses “the true kernel” of Popperian falsificationism. I suspect that (PP) is not so straightforward; what you should think of it depends on wider issues. Suppose  $H_1$  fails to entail not- $E$  because  $H_1$  says nothing whatever about  $E$ . Likelihoodists will then decline to say that  $E$  favors  $H_1$  over  $H_2$ ; they will say that if  $\Pr(E|H_1)$  makes no sense, then  $E$  does not favor  $H_1$  over  $H_2$ . Bayesians, of course, will not see things this way; if  $H_1$  says nothing about  $E$ , they will say that  $\Pr(E|H_1) = \Pr(E)$  and they will have no qualms about evaluating  $\Pr(E)$ . But likelihoodists have trouble with this quantity, the unconditional probability of the observations, because it is an average likelihood (Sober 2011, pp. 29–30):

$$\Pr(E) = \sum_i \Pr(E|H_i)\Pr(H_i).$$

The problem of priors rears its ugly head and so does the problem of having to survey the likelihoods of all possible hypotheses, or so likelihoodists will say.

### 3.3. *The Weak Law of Likelihood*

Fitelson likes the principle about favoring that Joyce (2003) calls the Weak Law of Likelihood:

(WLoL) If  $\Pr(O|H_1) > \Pr(O|H_2)$  and  $\Pr(O|\text{not-}H_1) \leq \Pr(O|\text{not-}H_2)$ , then  $O$  favors  $H_1$  over  $H_2$ .

Fitelson notes that the (WLoL) talks about “catch-alls” and so its antecedent will be controversial to “some philosophers.” He is right. In particular, *likelihoodists* dislike the (WLoL) for precisely that reason. Although it often makes sense to talk about the probabilities that scientific theories confer on observations, it often does not make sense to talk about the probabilities that the negations of those theories confer. Relativity theory says something about how probable Eddington's eclipse data was. What does the negation of relativity theory say about those observations? The negation is a very long disjunction of all

possible alternatives to relativity theory. The likelihood of a catch-all is a weighted average:

$$\Pr(O|\text{not}H) = \sum_i \Pr(O|H_i)\Pr(H_i|\text{not}H).$$

The first product term on the right-hand side takes account of *all* theories compatible with not-H—even those that have not yet been dreamed up. The second product term is something that also makes likelihoodists queasy. What is the probability of relativity theory, given that Newtonian theory is false? The likelihoods of catch-all are things that likelihoodists often won't touch with a stick. Fitelson's logically correct comment that the Law of Likelihood (LoL) entails the Weak Law of Likelihood (WLoL) needs to be understood in this light. Likelihoodists are not making a mistake when they endorse the (LoL) but think that the antecedent of the (WLoL) is often impossible to evaluate.\*

### References

- Durrett, R. (2002): *Probability Models for DNA Sequence Evolution*. New York: Springer.
- Fitelson, B. (2007): "Likelihoodism, Bayesianism, and Relational Confirmation." *Synthese* 156: 473–489.
- Joyce, J. (2003): "Bayes' Theorem." Stanford Encyclopedia of Philosophy. URL = <http://plato.stanford.edu/entries/bayes-theorem/>.
- Royall, R. (1997): *Statistical Evidence—a Likelihood Paradigm*. Boca Raton, Florida: Chapman and Hall.
- Sober, E. (1990): *Core Questions in Philosophy—a Text with Readings*. New York: Macmillan.
- (1993): *Philosophy of Biology*. Boulder, Colorado: Westview Press.
- (2004): "Likelihood, Model Selection, and the Duhem-Quine Problem." *Journal of Philosophy* 101: 1–22.
- (2008): *Evidence and Evolution—the Logic Behind the Science*. Cambridge: Cambridge University Press.

---

\* My thanks to Mike Steel for useful discussion.