

Critical Notice

ELLIOTT SOBER, *Evidence and Evolution: The Logic behind the Science*. Cambridge: Cambridge University Press, 2008.

INGO BRIGANDT

Department of Philosophy

University of Alberta

Edmonton, AB T6G2E7

Canada

As Elliott Sober acknowledges in the preface, the title of his latest book *Evidence and Evolution* is potentially confusing. For his discussion does not present various known empirical facts that support the theory of common ancestry, such as fossil data and genetic and anatomical features of extant species. Rather, it is an epistemological study of *scientific inference* and the confirmation of hypotheses in the context of evolutionary biology. Sober targets a dual audience. Philosophers will obtain from his detailed and convincing account general lessons on the nature of confirmation and an introduction to or overview of basic models in evolutionary theory. Biologists will benefit as well, since Sober steps back from the variety of existing quantitative models and identifies common strands in hypothesis testing and addresses how to justify certain fundamental assumptions that biologists often take for granted. Below I will highlight two philosophical themes that run through Sober's discussion. Sober has made these point more than once in previous work (Sober 1988, 1999, 2007), but they are of central importance and cannot be repeated often enough. These are the facts that (1) confirmation in science is contrastive, where a theory can be meaningfully tested only against one or more rival hypotheses, and (2)

scientific inference methods are not *a priori* and valid for every case, but presuppose specific empirical assumptions and thus are legitimate only in a certain range of empirical cases.

Apart from a conclusion, the book is divided into four long chapters (each of which is organized into several sections). Chapter 1 discusses different schools of hypothesis testing, including Bayesianism, likelihoodism, and frequentism. Chapter 2 addresses how the intelligent design hypothesis fares against evolutionary theory. Biologists especially may wonder what the point is of discussing a hypothesis without scientific merit. Yet Sober's account has implications for the confirmation of evolutionary theory and the testing of scientific theories in general, so that some of his major philosophical moves are made in this context. The subsequent chapter discusses how the tenet that a biological character is the result of natural selection can be tested against alternatives, such as being the outcome of random drift. Chapter 4 concerns common ancestry. While most biological accounts test different phylogenetic trees against each other, Sober focuses on how the very hypothesis that extant species have a common ancestry rather than independent origins can be confirmed.

To understand Sober's main thesis, let us take a look at Chapter 1, which offers a general discussion of the impact of evidence by comparing Bayesianism, likelihoodism, and frequentism. These approaches are not to be understood as different semantic or metaphysical interpretations of probabilities, but as different epistemological methods of scientific inference. *Bayesianism* offers an account of to what degree a hypothesis H should be believed, and how to change this degree of belief in light of new observational evidence O in accordance with Bayes's theorem:

$$P(H | O) = \frac{P(O|H)}{P(O)} \cdot P(H). \text{ One problem with Bayesianism is that in many concrete cases it is}$$

impossible to assign a non-arbitrary prior probability / degree of belief P(H) (from which to objectively update the degree of belief based on further evidence). *Likelihoodism* has the

advantage that it can do without subjective priors, as it focuses on the likelihood $P(O | H)$, i.e., how probable a certain observation O is according to hypothesis H . Likelihoodism gets traction by comparing the likelihoods of two or more hypotheses. If $P(O | H_1) > P(O | H_2)$, then observation O favours H_1 over H_2 ; and among a set of rival hypotheses, there may be one that is favoured over all others. However, while Bayesianism—if it is applicable—offers an answer to the question of which hypothesis to believe (at least to which degree), likelihoodism as such does not tell us what we should believe, it rather tells us what the evidence says (vis-à-vis one or several hypotheses, p.32).

Frequentism is not one inference method, but encompasses several different methods. One well-known such statistical tool is Fisher's significance test, where a null hypothesis is rejected if the p-value of an observation is below a chosen the level of significance, e.g., 0.05.¹ One problem with this widely used method is that—in contrast to Bayesianism and likelihoodism—it does not follow the principle of total evidence, which maintains that in assessing a hypothesis one should take all observational evidence into account (p.53). Another frequentist method is Neyman-Pearson hypothesis testing, which has the drawback that it violates the principle that if acquiring observation O justifies rejecting H , where this was not justified before O was obtained, then O must be evidence *against* H (p.58). To illustrate how these two frequentist inference methods compare to the other two approaches, Sober discusses stopping rules, i.e., the use of a rule that determines at what point of collecting observations the inquiry is over. For instance, when tossing a coin one may decide to stop after 20 tosses, and in one run it may turn out that one obtains exactly 6 heads. The same outcome (6 heads and 14 tails) may also be obtained if a

¹ The p-value is the probability of obtaining a result at least as extreme as the observed result, given the null-hypothesis.

different stopping rule is used, e.g., stopping after the 6th head has come up. Bayesianism and likelihoodism interpret the observed outcome independently of which stopping rule was used. Fisher significance testing and Neyman-Pearson hypothesis testing, however, may draw different inferences from one and the same observed outcome, depending on which stopping rule was used in generating the data (p.72).

The variety of frequentism favoured by Sober is model-selection theory, using the Akaike information criterion. This approach determines which model (hypothesis) in a class of models fits given observational data best. This method has often been misunderstood when pointing out that it may fail to converge on the true model (if the set of data points goes toward infinity). As Sober makes plain, this fact cannot be used as an objection, because model-selection theory is not meant to assess which model is true. Rather, it estimates the predictive accuracy of a certain model. Thus, model-selection theory tells us neither what hypothesis to take as true (as Bayesianism does) nor what the evidence says in relation to hypotheses (as likelihoodism does)—instead it addresses a third type of question about scientific inference, the goal of assessing a hypothesis's predictive accuracy (p.91). Unlike some frequentist methods, model-selection theory is essentially contrastive (similar to likelihoodism), since a certain model does not have an absolute score of predictive accuracy, but only a score relative to other models considered.

In addition to offering a good introduction to Bayesianism, likelihoodism, and frequentism, which will be helpful for novices, Sober offers many extra details and considerations on these methods. He also has general philosophical points in stock. Realism and instrumentalism have usually been cast as global hypotheses—one takes either one or the other stance about scientific models and theories. The model-selection approach suggests that this is a mistake. Sober argues that the Akaike framework yields a “mixed philosophy” (p.98): an *instrumentalism* about models

(models are predictively accurate but not taken as true), but a *realism* about fitted models (closeness to truth can be measured for fitted models). Given the existence of different schools of hypotheses testing, Sober announces that “no single unified account of scientific inference will be defended here” (p.3). “I am prepared to be a Bayesian on Monday, Wednesday, and Friday, a likelihoodist on Tuesday, Thursday, and Saturday, and a model selectionist on Sunday” (p.356). This main tenet of *Evidence and Evolution* is backed up in the subsequent chapters, where Sober shows in concrete situations in evolutionary biology when Bayesianism is the best option (if priors and likelihoods are available), when likelihoodism is called for (if likelihoods but no priors are defensible), and when and why model selection is the right inference method. As a result, Sober is a *pluralist about inference method*. But he is at pains to argue that this does not imply a pluralism about the concept of evidence (p.357). For different inference approaches address different questions: What should I believe? What does the evidence say? Which hypothesis has the highest predictive accuracy? And it is not surprising that different questions have different answers, and that different methods are used in different epistemic contexts.

Sober is right that using Bayesianism, likelihoodism, and model selection does not lead to several notions of evidence, and that there are different epistemological questions or inferential aims. However, Bayesians also view themselves as addressing the likelihoodist question of what the evidence says. There are also cases where $P(O | H_1) > P(O | H_2)$, yet against likelihoodism observation O ought not to count as favouring H_1 over H_2 , as O logically entails H_2 , but does not entail H_1 (Fitelson 2007). To avoid this conflict of inference methods, Sober must view likelihoodism as a fallback position that is to be used only if Bayesianism cannot be used. Even if different methods can be understood so that philosophically speaking they are orthogonal, they may lead to conflicts in scientific practice, as scientists understand them as guidelines about which hypothesis to (tentatively) accept and reject. For instance, even if properly interpreted

likelihoodism does not answer the question of whether H_1 is to be ‘believed’ to a higher degree than H_2 , but answers whether H_1 is ‘favoured’ over H_2 , in scientific practice this amounts to the same. If both Bayesianism and likelihoodism are applicable to a range of hypotheses in the face of evidence, the former may lead to one hypothesis being chosen (as it has the highest posterior probability), and the latter may result in a different hypothesis being selected (as it has the highest likelihood).

I The Testability of Intelligent Design

While some may feel that a discussion of the epistemic credentials of intelligent design is moot (and be inclined to jump to the chapters on the testability of various evolutionary hypotheses), Chapter 2 actually presents some of Sober’s most insightful ideas on testability—especially the notion that confirmation is always contrastive. Of course scientists often happen to compare several rival hypothesis, but Sober makes the normative point that sometimes a hypothesis can be said to be testable *only* with respect to another hypothesis. These basic ideas have already been introduced by Sober in previous writings (Sober 1999, 2007), but they deserve to be repeated due to their philosophical importance and the fact that most critiques of intelligent design have not been aware of one of the most central fallacies underlying most intelligent design arguments.

Intelligent design is the idea that organisms could not have originated by natural processes alone, but only by the direct influence of some supernatural designer. (Theistic evolutionists, in contrast, assume that God created natural laws and processes, which in turn led to organisms by purely natural means.) A common argument for intelligent design is actually an argument against evolutionary theory. The idea is that the evolution of complex features (be it morphological structures, be it genetic or biological information) by means of random mutation and natural

selection is so improbable that one may take evolutionary theory to be false. Such probability arguments have been raised many times in various disguises, but have been developed prominently by the mathematically trained theologian William Dembski (1998, 2002), whose work has been repeatedly been scrutinized by Sober and others (Elsberry and Shallit in press; Fitelson et al. 1999). Dembski argued that an event whose probability is lower than his universal probability boundary of 10^{-150} cannot have resulted by naturalistic, non-intelligent means.² Even if it is pointed out that the improbability or even falsity of evolution does not entail the truth of intelligent design, the improbability of evolution alone would indeed be a devastating result. Yet it is based on a profound but extremely common fallacy. Given a hypothesis H (say, evolutionary theory) and an observation O (the presence of complex structures), it is argued that O would be extremely unlikely assuming theory H, that is, $P(O | H) \approx 0$. This is true, but the intended conclusion is that evolution is improbable given the observed complexity, that is, $P(H | O) \approx 0$. But the likelihood $P(O | H)$ being very small does not entail the posterior probability $P(H | O)$ being small. In fact, the posterior can have *any* value, if the prior $P(H)$ and expectedness $P(O)$ have appropriate values, as shown by Bayes's theorem: $P(H | O) = \frac{P(O | H)}{P(O)} \cdot P(H)$. As a result, *nothing* about the probability or improbability of evolutionary theory (given the presence of complex structures) can be concluded from the fact that the presence of complex structures would be extremely unlikely if evolution had occurred.

Sober offers a deeper analysis of this issue. The fallacy is usually not recognized as such (and even made by Fisher's statistical significance test), as it can be understood as a probabilistic

² Dembski obtained this number by multiplying the number of particles in the known universe, the maximal rate of change in physical states, and the age of the universe, multiplying again with one billion.

version of *modus tollens*. Recall that the above (fallacious) argument against evolution proceeds from the fact that $P(O | H) \approx 0$, which is equivalent to $P(\neg O | H) \approx 1$. According to *modus tollens*, from $H \rightarrow \neg O$ and O one may validly infer that $\neg H$. In a similar vein, the probability argument against evolution starts out with $P(\neg O | H) \approx 1$ and O , and wants to infer that $P(\neg H) \approx 1$, that is, $P(H) \approx 0$. Yet as Sober is at pains to argue, there is simply no probabilistic analog of *modus ponens* (p.130).³ It is actually easy to see why. Extremely unlikely events happen all the time. If one tosses a fair coin often enough or considers the conjunction of a sufficient number of probabilistic occurrences, one can create an event whose probability is arbitrarily small. Theory H may at the same time claim that $P(O_1 | H) = 1/2$ (a single coin toss), and that $P(O_2 | H) \approx 0$ (many tosses of the coin)—and both a true and a false theory can claim such probabilities. Thus, small probabilities in no way permit an inference about the probable falsity of the theory postulating this probability. The problem with Dembski's universal probability boundary is not that it is still too high a probability, but that any small probability value alone is *irrelevant* and does not even permit some inductive inference about the probability of evolutionary theory (p.51).

Probabilities (in the form of likelihoods) can matter, but only if *two or more hypotheses* are compared. If $P(O | H_1) = p_1$ and $P(O | H_2) = p_2$, then likelihoodism states that observation O favours H_1 over H_2 if and only if $p_1 > p_2$. The latter— $P(O | H_1) > P(O | H_2)$ —does not entail any particular relation between the Bayesian posterior probabilities $P(H_1 | O)$ and $P(H_2 | O)$, as long as no prior

³ Sober's recent book introduces this fact in Chapter 1, in the context of frequentism, as one problem with Fisher's significance tests is that they are an instance of probabilistic *modus tollens* (p.53). Yet in Chapter 2 more interesting conclusions are drawn from it. My presentation of this chapter highlights this issue much more than Sober does, given how widespread the fallacy from small probabilities is and how prominent it has been in a variety of intelligent design arguments.

probabilities $P(H_1)$ and $P(H_2)$ are assigned—which would be question-begging or at least contentious in the case of hypotheses such as evolution or intelligent design. But even the use of a likelihoodist framework permits some genuine scientific inference. Once different hypotheses and different likelihoods are compared, such inferences do not rely on a fallacious probabilistic analog of *modus tollens*. Now the question is what probabilities different hypotheses assign to a possible observation. This makes it necessary to get clear about what a certain intelligent design hypothesis and rival theories actually assert. Before Darwin, for instance in the historically prominent intelligent design argument of William Paley, the main alternative to intelligent design was Epicureanism, i.e., the idea that orderly physical and functional biological features have resulted by pure chance from physical particles whirling at random in the void. From this perspective, the question is whether $P(O \mid \text{Intelligent design}) > P(O \mid \text{Chance})$, where O are observed biological facts.

One possible position is to grant that the evidence favours intelligent design over pure chance, while pointing out that since Darwin we have another relevant hypothesis, and that $P(O \mid \text{Darwinian evolution}) > P(O \mid \text{Intelligent design}) > P(O \mid \text{Chance})$. Indeed, many contemporary creationists and intelligent design proponents dishonestly obscure this point by assimilating evolution to pure chance, e.g., when claiming that evolution is as improbable as a hurricane assembling scattered pieces of metal into a functioning airplane when blowing through a junkyard. Sometimes arguments against evolution simply assert that the evolution of complex features is extremely improbable, without offering any calculation of an actual probability value. Often some probability is derived, but strongly misrepresents what probability would be assigned by evolutionary theory, e.g., by pretending that evolution merely consists in the random change of a single genome, which ignores that mutation and reproduction occur in hundreds of thousands of individuals at the same time, and that because of natural selection only the more successful

gene variants are represented in the next generation (so that later generations do not start from a random and arbitrary starting point).

Interestingly, Sober does *not* favour the above reply that grants that $P(O \mid \text{Intelligent design}) > P(O \mid \text{Chance})$ but argues that $P(O \mid \text{Darwinian evolution}) > P(O \mid \text{Intelligent design})$. For “there is a devastating objection to Paley’s argument that does not depend in any way on Darwin’s theory” (p.126), as $P(O \mid \text{Intelligent design})$ need not be defined at all. Contemporary intelligent design proponents try to hide the religious nature of their approach, and thus often define intelligent design as the hypothesis that some intelligence somehow influenced the history of life at some point. This hypothesis does not predict any observable feature, not even in a probabilistic fashion. If $P(O \mid \text{Intelligent design})$ is indeterminate, this hypothesis is untestable, even untestable relative to other hypotheses. By attempting to portray intelligent design as a scientific theory, its proponents have rendered it untestable.⁴ As Sober points out, intelligent design can be tested if it is combined with *auxiliary assumptions*. If it is specified what the goals and the abilities of the designer are, probabilities can be assigned to predicted observations. This shows that even in a probabilistic framework (where predicted observations are not logically deduced, but assigned probabilities), many hypotheses can be tested only when coupled with auxiliary hypotheses.

Paley took some aspects of the designer’s goals and abilities for granted, so as to estimate $P(O \mid \text{Intelligent design})$ to be fairly high—at least higher than the chance alternative. Yet rival hypotheses such as chance and evolution do not make and in fact disagree with these auxiliary assumptions, so that intelligent design faces the need to offer *independent evidence* to support these assumptions—evidence that is independent of the very intelligent design hypothesis and of

⁴ Sober does not explicitly make this last point, but I found it worth highlighting.

the observation used to test intelligent design (e.g. well-adapted biological features). Needless to say, there does not seem to be any way to adduce such independent evidence that would justify claims about a supernatural designer's goals and abilities, creating an insurmountable problem for turning intelligent design into a testable theory without making question-begging assumptions. Paley famously illustrated the intelligent design argument by a pocket watch found on a heath (Paley 1802). Given the functional arrangement of the watch, one may infer that a designer has made it. Likewise, functional and complex structures like the eye are claimed to permit an inference to a designer of organisms. The design argument applied to watches is sound, however, it does not hold for biological structures as only in the former case the intentions of the designer (a human watchmaker) are known. Contemporary intelligent design proponents claim that they can infer a (supernatural) designer of biological systems just like intelligent agency can be inferred in archaeology and forensic science. Archaeologists can indeed infer for an object (e.g. a stone) whether it is more likely a human artefact—but only to the extent to which there is some information about how humans as opposed to natural processes shape natural objects. Forensic science can determine whether a fire was due to arson, given knowledge about how arsonists and other, accidental causes of fire proceed—information, which is not available for a hypothesis about supernatural designers.

As Sober makes plain, this consideration cuts both ways: if intelligent design cannot be effectively confirmed by observations, it cannot be *dis*confirmed by evidence. A common objection to intelligent design proceeds from the existence of maladaptive features such as the human appendix or panda's thumb, which could have been much more effectively designed by an independent creation of the species. (Recently this has been put in terms of the idea that 'no designer worth his salt' would create so many suboptimal and maladaptive features.) Claiming that such observations are very unlikely according to intelligent design would likewise be

possible only if one had evidence about the designer's intentions—but the critics of intelligent design do not have this information either (p.128). Sober makes a similar point regarding the argument from evil, in which case it may be possible but difficult to support auxiliary assumptions that entail that the existence of evil favours atheism over theism (p.167). Sober is right about this, but he overlooks the possibility that a proponent of evolution can for the purpose of argument accept the auxiliary assumptions made by the *intelligent design proponents*, and then argue that the observed biological features—many of which are suboptimal—favour evolutionary over intelligent design theory. This is in line with Sober's endorsement of the *principle of total evidence*, the epistemological principle that in hypothesis testing one must take all observational evidence into account. As he points out, Paley cherry-picked when focusing on the functionality of a watch and the adaptive features of biological systems, and using this as the evidence for intelligent design (p.134). Instead, one has to consider all features of a system when using these as the evidential basis to compare chance, intelligent design, and evolution—not only the adaptive or just the maladaptive traits.

In addition to the possibility of testing intelligent design in a likelihoodist framework—which proceeds from $P(O \mid \text{Intelligent design})$ —Sober discusses whether it is possible to estimate $P(\text{Intelligent design} \mid O)$ by construing the intelligent design argument as an inductive sampling argument. The posterior probability $P(O \mid \text{Intelligent design})$ would make a direct verdict about the degree of confirmation or the probable truth of intelligent design, yet Sober argues that this particular kind of argument is invalid (p.175). He also critically addresses Michael Behe's argument for intelligent design based on the notion of irreducible complexity (p.155). This discussion is fairly brief and overlaps with the extensive earlier critiques of Behe by other authors. Overall, apart from the need to adduce (independently supported) auxiliary hypotheses, the main philosophical upshot of this chapter is that a hypothesis may be testable only relative to

other hypotheses. An earlier example of Sober's even shows that a hypothesis can be testable relative to one alternative, but untestable relative to another alternative. The claim that there is an odd number of letters on a certain page of a book is testable against the claim that the page contains an even number of letters. However, the former is not testable against the hypothesis that there is a demon manipulating us such that we falsely believe that there is an odd number of letters on the page (Sober 1999).

Throughout the book Sober emphasizes that many hypotheses only make testable predictions in combination with underlying empirical assumptions (p.333). It has to be made explicit that these presuppositions are made; and, as they are not *a priori*, they are in need of empirical defence. Under the label of the Duhem-Quine thesis, this fact has been used to draw strong philosophical conclusions. If the hypothesis H to be tested alone does not predict observation O, but only in conjunction with auxiliary assumption A, then the claim is that evidence O cannot test H specifically, but tests the conjunct H&A. Quine (1953) boldly argued for an *epistemic holism*, according to which evidence bears not on individual beliefs, but on our total set of beliefs. While it is not a focus of *Evidence and Evolution*, in an earlier discussion on testability Sober (1999, p.54) used his ideas to reject epistemic holism. Sober pointed out that while observations are theory-dependent, they may not be dependent on the theory under test. If two rival hypotheses tested against each other agree on certain theoretical assumptions, the latter do not raise epistemological problems for observations that depend on them. Sober rejected epistemic holism on the grounds that there is often *independent evidence* for auxiliary assumptions. Given prior independent support for A, novel evidence O tests hypothesis H (and not just the conjunct H&A). Scientists "don't simply *invent* auxiliary assumptions ... the auxiliary assumptions *used* in a test and the hypotheses *under* test differ in their epistemological standing" (1999, p.54).

I agree with Sober on this point, but some clarifications are necessary for a philosophically

legitimate construal of his important insights.⁵ If one's epistemological project is of an ambitious *foundationalist* nature, in the sense of offering an absolute justification for any and thus every warranted belief, then some sort of epistemic holism is unavoidable.⁶ For in the case of independent evidence E supporting auxiliary hypothesis A, one may apply the same argument to this hypothesis. Auxiliary A entails observation E only in conjunction with further auxiliary assumptions, so that technically speaking E cannot support auxiliary hypothesis A taken in isolation, undermining Sober's idea that observation O predicted by H&A tests H due to A being independently supported. However, if instead one's philosophical project is an *epistemology of science in practice*—which is in my view the more relevant agenda—then Sober's arguments succeed and his ideas have philosophical impact. In their actual practice, scientists never attempt to (and do not pretend to) have full evidential support for all of their scientific beliefs. The latter are genuine knowledge from the point of view of an epistemology of science in practice, though a very traditional, foundationalist epistemology's construal of 'knowledge' may not apply. While an auxiliary assumption in isolation cannot be supported by a specific body of (theory-independent) evidence, Sober is right that scientists have *some* evidence for the auxiliary

⁵ I benefit from discussions with Jordan Glass on this issue.

⁶ My notion of a foundationalist *endeavour* is not identical with foundationalism—as opposed to coherentism—in epistemology, which is a *doctrine* about the structure of knowledge systems. Still, the simplistic type of foundationalism that Quine (1953) calls 'reductionism' (each empirical statements corresponds to and thus can be fully justified by some sense data or observation statements) can be motivated by what I call a foundationalist endeavour. The foundationalist endeavour may seem similar to the 'first philosophy' that Quine (1969) rejects, yet while first philosophy construes philosophy as distinct from science, there can be a foundationalist endeavour that aims at offering a justification of belief using scientific knowledge. We shall see that I view even such a foundationalist project as misguided, and Quine as being influenced by a foundationalist endeavour despite his naturalistic rejection of first philosophy.

assumptions that have to be made to test a target hypothesis. Rather than testing and empirically supporting *every* of their beliefs, scientists focus on certain questions that are currently *relevant* or on certain hypotheses that are currently controversial. Given this, when evaluating hypothesis H_1 against H_2 , where the test requires assumption A on which both hypotheses agree, it is perfectly legitimate for scientists to not worry about whether A is sufficiently supported by independent evidence. At a later point (e.g. when scientific questions or assumptions change), assumption A may come under scrutiny, and only then more evidence has to be put forward to back it up or it has to be discarded.

Quine was right that *in principle* any empirical belief can be held on to (in the face of any recalcitrant evidence) if other beliefs are suitably modified, and that in principle a given observation may bear on any theoretical belief. But such an epistemic holism ignores the epistemic structure and dynamic nature of science where at any *actual* point in time only some questions are scientifically relevant and only some assumptions are considered in need of defence. Given any particular situation in science, epistemic holism does not obtain, as either there is independent evidence for auxiliary assumptions that is sufficient given the *current* knowledge and problem situation, or all *currently* viable rival hypotheses agree on the auxiliary assumptions (so that the hypothesis, but not the auxiliary is genuinely tested). This situation will change, but at any point in time only some beliefs are under scrutiny while others may be taken for granted. An epistemic holism obtains only if one is still wedded to an outdated foundationalist project. An epistemology of science in practice does not endeavour to offer an absolute justification for any warranted belief, but rather focuses on how a belief system can be and is epistemically improved. Even though one can use Quine's web of belief metaphor to motivate a shift from belief justification to rational belief revision, and even though Quine (1969) rejected 'first philosophy' and proposed a naturalistic epistemology, his focus on logical theory structure

and his *in principle* arguments about evidential relations show that he still thought in terms of some foundationalist project, which from the point of view of an epistemology of science in practice fails to address what philosophically matters about understanding actual science and its rationality.

An important insight of Sober's is that confirmation is contrastive, where two or more rival hypotheses are compared. One complication that he does not mention is the issue of explanatory loss, where a new theory gains significant support from being able to account for a phenomenon that was unintelligible for earlier approaches, yet the new theory (at least in its initial formulation) cannot account for everything that an older theory could. In such a situation the two theories (even though they are about the same domain) are best suited for different observations, and it may be impossible to obtain a verdict as to how well each of them accounts for the *totality* of observations, so as arrive at an effective hypothesis comparison, e.g., $P(\text{total O} \mid H_{\text{new}}) > P(\text{total O} \mid H_{\text{old}})$. Moreover, apart from quantitative relations between a hypothesis and some observations (such as a likelihood that represent to degree to which the hypothesis *predicts* the observation), scientists use a variety of other considerations to assess the epistemic status of theories, for example, how unified one theory is, or to which extent it offers a causal-mechanistic *explanation* of the phenomena in its domain.

II Testing Claims about Natural Selection

So far Sober has referred to testing 'evolutionary theory' (against intelligent design). However, evolutionary theory is not a single model or hypothesis at all, and what evolutionary biologists do in practice is to test rival evolutionary scenarios against each other. This is the topic of the rest of Sober's book. Chapter 3 is about testing hypotheses claiming that natural selection has influenced

a trait's evolution. For instance, given a quantitative character able to undergo continuous variation, one may wonder whether the (average) character state observed in an extant species is simply due to random genetic drift, or whether natural selection has played a role apart from drift. To test two such models against each other, Sober adopts a likelihoodist framework (as no meaningful prior probabilities can be assigned to each hypothesis), and compares the likelihood of the present state given drift only and given drift plus selection, based on such assumptions as what the ancestral state and the strength of selection and drift were. Which of the two hypotheses is favoured nearly always depends on these additional assumptions specific to a concrete case. In fact, one result that will strike many novices as counterintuitive is that even if since its ancestral state the character has evolved *away* from the optimal state favoured by natural selection, there are conditions (e.g. long evolutionary time spans) where the selection plus drift hypothesis is favoured over the pure drift model (p.201). As in intelligent design critiques of evolution, two or more hypotheses and their likelihood values have to be compared. A probabilistic version of *modus ponens* is invalid, so that the fact that an observed trait value would be extremely unlikely assuming that no selection has taken place is no argument against this hypothesis of pure drift.

The adoption of such a *likelihoodist method* cannot be defended *a priori*; instead, it is advantageous in and applicable to certain kinds of cases. The models tested against each other make certain empirical assumptions, e.g., about what the ancestral state was and what trait value is favoured by selection. Sober explains how independent evidence can be used in attempts to back up such assumptions. However, this often turns out to be difficult, so another strategy to increase inferential power—in the sense of arriving at better supported hypotheses—is to *change the explanandum*, focusing instead on a different kind of hypothesis that is weaker but still of substantial scientific importance. Rather than wondering why a particular species has a certain character value, one may inquire what accounts for the correlation of two characters exhibited

across several species, e.g., why bears living in colder climates tend to have longer fur length. In this case, to assess selection versus drift hypotheses one need not know about the ancestral and the optimal character value, the heritability of the character, or the population size. The basic reason is that causes are difference makers, and the operation of one kind of causal process can be ascertained even if other causal factors affecting each individual species are unknown. In this context Sober also addresses *Reichenbach's principle of the common cause*, which maintains that if two variables are correlated, then unless one is the cause of the other, both are the effects of a common cause. He argues that if the likelihood approach is right, then Reichenbach's principle is too strong. For according to the former one cannot even conclude that a correlation was *probably* due to a causal connection; rather, one may conclude that a causal connection hypothesis is better supported by correlational data than a causal independence hypothesis (p.231).

Given correlations and variation patterns across several species, one can use model-selection theory—as a different, frequentist inference method—to assess how well different explanatory hypotheses that differ in their degree of *unification* fit the observed data. For instance, for a given trait one can assess whether the trait's average value in humans is explained (in a unifying way) by the same factor that explains the character's variation among other species, and/or whether the trait's variation in humans is explained (in a further unifying way) by the same factor that explains the character's variation among other species (p.228). A hypothesis that explains correlations can do with fewer empirical assumptions than a hypothesis accounting for a particular trait value in a species, but the former still has to make (and defend based on independent evidence) an assumption about the slope of an optimality line (fitness curve). Several approaches using molecular data test selection versus drift claims with even less auxiliary assumptions. One possibility is to consider DNA sequences in at least three different species and perform a relative rate test, which compares the rate of molecular evolution in

different lineages. Sober argues that such hypotheses are best inferred from data using a *model-selection approach*, which in this context is more defensible than Bayesianism or Neyman-Pearson hypothesis testing (p.239).

Model-selection theory is also useful in understanding the notion of *parsimony* when it comes to testing natural selection versus phylogenetic inertia. Traits in two or more species can be similar because they face similar selection pressure (the traits have similar optimal values). But of course the similarity can be due to the species having a common ancestor and the value of the ancestor's trait not having changed much in the lineages leading up to the extant species. Biologists call this latter causal influence phylogenetic inertia (though Sober points out that this term is confusing and be better dubbed 'ancestral influence'). A widespread assumption is that since all species are evolutionarily related, phylogenetic inertia always plays some role and selection should be *additionally* invoked only if phylogenetic inertia alone cannot explain the data. The idea is that assuming mere phylogenetic inertia is more parsimonious; and parsimony arguments appear in different inference contexts in evolutionary biology—being nothing but an application of *Occam's razor*. However, Sober rightly points out that the more parsimonious hypothesis can neither be simply taken as the default model, nor empirically accepted because it is more parsimonious. Instead, different models have to be actually compared to each other. Each embodies underlying empirical assumptions about causal factors and historical processes, which must be considered. A comparison of a phylogenetic inertia and a selection hypothesis may yield that both are able to account for the data, and that given this evidence neither can be preferred over the other—even if one is more 'parsimonious' (p.250). As a result, Sober is sceptical of Occam's razor as a principle that holds *a priori* and in every empirical context.

This chapter offers some points that are of general interest to philosophers, and it helps biologists step back from the flurry of models that exist in evolutionary theory and see some

basic principles of evolutionary hypothesis testing. However, given that Sober can address only some basic approaches, he could have offered some account of how the material covered articulates with more advanced models, at least by pointing to chunks of the biological literature that his discussion can be viewed as reaching out to. For instance, what about more complex models that combine both phylogenetic inertia and selection? In a similar vein, *adaptationism*—the idea that virtually any trait is to be treated as an evolutionary adaptation and thus the product of natural selection—has been a major point of contention within evolutionary biology. Some biologists have claimed that when an explanation that claims a trait to be an adaptation for one selectively advantageous function turns out to be untenable, then another adaptation explanation is offered, and so on, but the idea of adaptationism is never rejected and thus never tested or empirically supported (Godfrey-Smith 2001; Gould and Lewontin 1979). Philosophers have prominently criticized alleged evolutionary explanations of human behaviour, and more generally scrutinized whether adaptationism as found in sociobiology and evolutionary psychology meets scientific standards (Buller 2005; Kitcher 1985; Richardson 2007). Curiously, Sober does not address adaptationism, even though he covers the testing of selection hypotheses. So an account of how his chapter bears on the adaptationism debate would seem mandatory (even given that his earlier writings addressed adaptationism, e.g. Sober 1993). In their actual explanations of the evolution of concrete anatomical characters, many biologists avoid illicit versions of adaptationism by using a detailed comparative and phylogenetic study that among other things informs them about what characters are separate units that can be subject to selection and evolve independently of other characters. Sober could have indicated what important evidential considerations and explanatory practices found in current evolutionary biology complement the selection vs. drift considerations he covers.

III Testing Claims about Common Ancestry

Chapter 3 addresses how to test hypotheses about phylogenetic relationships. Sober has already extensively published on this issue; in fact, he is one of the few philosophers who have discussed inferring phylogenetic trees (Sober 1988). Given a certain set of extant species, biologists assess rival phylogenies, and there is a substantial literature on which methods are most reliable. The novelty of Sober's discussion is that rather than testing different concrete phylogenies—all of which assume that the species involved have a common ancestor—the focus is on how to test common ancestry in the first place. To this end, Sober compares the hypothesis that a certain number of species share a common ancestor with the alternative that they evolved independently of each other (even creationists assume that upon creation each species has had some evolutionary history). Actually, not only is this a philosophical consideration about the foundations of evolutionary theory, but it is germane to a currently debated biological issue. While most biologists assume that all living organisms are descended from one ancestral individual (the so-called Last Universal Common Ancestor), there are some suggesting that the three major groups of organisms—the eukaryotes, the archaea, and the bacteria—stem from a complex web of interacting single-celled organisms that cannot be resolved to a single ancestral organism (Doolittle and Baptiste 2007).⁷ Sober's discussion surrounds the basic inference principle that he dubs *modus Darwin*, i.e., the prominent idea that similarities among species are evidence for common ancestry. To assess the common ancestry versus the separate ancestry hypothesis, Sober rejects Bayesianism, as the prior probability of each hypothesis (independent of any evidence such as observed similarities) are empirically contested and very hard to estimate

⁷ Due to lateral gene transfer across primitive organisms from very different taxonomic groups, genetic data about extant organisms may forever be inadequate to settle which hypothesis is the correct one.

(p.277). Instead he views *modus Darwin* as a kind of likelihoodist inference, so that one has to determine which hypothesis the evidence favours, i.e., whether observed similarities are more likely given the common ancestry or given the separate ancestry hypothesis.

Dealing first with the most simple case, where the only evidence about the species considered is one character that can be in two states is, Sober sets up nine conditions that jointly are sufficient for common ancestry being favoured, but none of them is a necessary condition. The conditions describe various empirical assumptions in probabilistic terms, expressing e.g. the idea that two lineages stemming from a common ancestor evolve independently of each other. All these conditions are empirically quite likely so that *modus Darwin* should have a broad range of applicability. Since none of the conditions is necessary for common ancestry having a higher likelihood, they are not assumptions that *modus Darwin* requires. The interesting question is what happens if one of the conditions does not hold. In such a case, whether the common ancestry or the separate ancestry is favoured depends on which condition is not met. Thus, Sober rightly stresses that “there is no *intrinsic* reason why similarity must matter as evidence for common ancestry” (p.283). Both a similarity and a dissimilarity in two species is consistent with them having a common ancestor (different traits can be due to evolutionary divergence since the common ancestor), and both situations are consistent with the species having no common ancestor. Whether or not *modus Darwin* is a legitimate inference principle depends on empirical background assumptions—there are possible situations in which similarity even favours the separate ancestry hypothesis. Sober discusses various more complex models, such as characters that can have several or continuous states. In the case of several characters being considered, Sober makes plain that in general it is not the case that common ancestry is favoured if across two species there are more similar characters than there are dissimilar characters. Which hypothesis about common ancestry is favoured depends on various empirical considerations

about the evolutionary processes creating the final outcome, e.g., whether pure drift, frequency-independent selection, or frequency-dependent selection is at work, and which character state was favoured by selection (the ancestral or the derived state, the state shared by the majority or minority of extant species). Other considerations discussed by Sober are how such additional sources of evidence as intermediate fossils and the biogeographical distribution of species can be used in evolutionary inference.

One long section addresses a topic that aligns with Sober's earlier work (Sober 1988) and ties much more directly into issues that have been the subject of the biological literature. This is the comparison of *maximum likelihood* and *maximum parsimony* as two different basic approaches to phylogenetic inference, where some scientists prefer one over the other method. Parsimony is a prominent inference method that has been used for a few decades. It asserts that among rival phylogenetic trees, the one is to be preferred that requires the smallest total number of evolutionary changes (genetic mutations, or changes in character states, depending on whether molecular or anatomical data is used). A less parsimonious tree assumes that equivalent evolutionary changes have taken place independently of each other in two or more lineages (so that the total number of changes required becomes larger). Maximum likelihood is the likelihood inference method mentioned above—among rival phylogenies, the one that assigns the highest likelihood to the observed features is to be preferred. We saw that determining how likely certain traits in extant species are (given a certain phylogenetic hypothesis) is possible only once substantial assumptions are made about the evolutionary processes resulting in the observed pattern. This is well-known known from the biological literature; indeed, a standard criticism of maximum likelihood approaches is that they make empirical assumptions about historical factors whose truth simply cannot be ascertained. In the case of maximum parsimony, a common objection is that it unrealistically assumes that evolutionary change always proceeds in the most

parsimonious way. Sober offers several clarifications in this debate (p.334ff).

Evolution always proceeding in the most parsimonious fashion is a *sufficient* condition for maximum parsimony being a very reliable inference method. However, it is not a *necessary* condition, and thus the use of maximum parsimony does *not* make the implicit assumption that evolution is parsimonious. As far as maximum likelihood is concerned, Sober compares several well-known models that fall into this category, and differ in terms of how complex they are, where less simple and idealized models have more free parameters that permit them to embody more realistic assumptions about the evolutionary process. Even though these different models are usually compared using a Bayesian framework, Sober shows that also a frequentist, model-selection approach can shed light on their relative strength. When comparing maximum parsimony versus maximum likelihood, Sober considers two criteria of assessment. Ordinal equivalence is the situation that method A favours hypothesis H_1 over H_2 if and only if method B likewise favours H_1 over H_2 . It is known that parsimony and maximum likelihood are ordinally equivalent when the probabilities of change in character states are very small. This has often been taken to imply that a maximum parsimony approach assumes that evolutionary change is very improbable, but Sober makes plain that this is not the case (as it confuses sufficient and necessary conditions). Another evaluation criterion is statistical consistency, which is the property that an inferred hypothesis converges to the true situation if the size of the data points (the quantity of evidence) goes to infinity. As is well-known from the theoretical literature, there are possible cases where maximum parsimony is not statistically consistent (in fact there are scenarios such as substantially different rates of evolutionary change in less closely related lineages where maximum likelihood systematically converges to a false phylogenetic tree if the data set increases).

Though there are empirical cases where parsimony is reliable, this shows that parsimony is

not an *a priori* inference method independent of any empirical presuppositions; and thus despite its similarity to *Occam's razor* it is not to be taken as an application of this metaphysical principle to evolutionary hypotheses. In fact, the failure of maximum parsimony in certain cases (and thus its dependency on empirical assumptions) supports the case that *a priori* metaphysical principles such as Occam's razor are illegitimate, at least for assessing empirical hypotheses. In addition to this lesson for philosophers, Sober's point also bears on a strand within phylogenetic systematics that strives for systematics being autonomous and independent of evolutionary theory. This approach (called pattern cladism) erroneously claims that using maximum parsimony one can establish phylogenetic *patterns* (i.e., phylogenetic trees) independently of assumptions about the characteristics of the evolutionary *process* (merely assuming that evolution has occurred).

In a subsection entitled "Homology," Sober addresses the common idea that homologies among species provide evidence for common ancestry. However, given a definition of homologies as similarities *due to common ancestry*, "If our goal is to *test* the common-ancestry hypothesis against the separate ancestry hypothesis, then it would beg the question to say that our data consists of 'homologies' in this sense" (p.283). Sober is absolutely right on this point, but a careless reading of his remarks may erroneously suggest that he endorses the circular reasoning challenge that creationists have repeatedly levelled against evolution based on the notion of homology. The objection is that phylogenetic trees are inferred from homologies, but since homologies are defined as similarities due to common ancestry, homologies can only be inferred from phylogenetic trees, resulting in a circular argument. Biologists of course do not engage in this kind of circular reasoning. Instead, they start out with similarities among living species—data that can be observed without any phylogenetic assumptions—and from this establish a phylogenetic tree (the one that explains the total data best), and finally read off from the

phylogeny whether a similarity is due to common ancestry—and thus a homology—or due to independent evolution. Sober has indeed covered this method of how to infer phylogenetic trees in a legitimate fashion (the above quote is fully consistent with it and merely points out that the data feeding into phylogenies are not homologies but similarities, including those due to parallel evolution). But given the prominence of the ‘homology is circular reasoning’ objection, he should have addressed it explicitly and explained how his account can easily avoid it.

More importantly, while Sober is right that homologies in the sense of similarities due to common ancestry are not evidence for phylogenetic hypotheses, this conflates two notions of homology and actually ignores the historically more original understanding of homology and its evidential impact. The difference between these two notions can be made plain by using the distinction between a character and a character state (Brigandt 2007). Properly construed, a *character* is a bodily part found in several species (e.g. the thighbone), which can vary in its internal properties across these species, and thus can be in different *character states* in different species (the particular length and shape of the thighbone in humans, horses, and pigeons). A homology as used by Sober in the sense of a similarity (due to common ancestry) refers to similar *character states* in different species.⁸ However, as originally introduced, homology is a principle of morphological organization, namely, the fact the same bodily unit occurs in different species, even though this morphological unit can vary across these species. Before the advent of Darwinian evolutionary theory, Richard Owen defined a homologue as the “same organ in different animals under every variety of form and function” (Owen 1843, p.379). Thus, a homologue is the same character in different species. Homology in this sense refers to the

⁸ In the biological literature this is called a taxic construal of homology, which equates the word ‘homology’ with the notion of synapomorphy.

identity of characters across species (bodily units identical in type), not to be confused with *similarities* that homologous characters exhibit in more closely related species. Given evolutionary change, even dissimilar traits can be the same, homologous character. Already pre-Darwinian comparative anatomists had been able to reliably establish homologies (*sensu* identical characters) across species, and the fact that characters in taxonomically unrelated species (e.g. amphibians and mammals, and even fish and mammals) had come to be recognized as homologous proved to be an important reason for why Darwin's theory was accepted by the biological community soon after having been proposed. For common ancestry provides an explanation for why the same character (homologue) occurs in many different species. Given that many such homologies had been established before the advent of evolutionary theory and based on criteria that do not presuppose evolution, homology in the sense of identical characters provides a *non-circular* argument in support of common ancestry. Sober fails to address this point and to discuss its inferential logic, as his construal of 'homology' is restricted to homology in the sense of similar character states (he seems to be oblivious of the more traditional and still prevalent notion).

IV A Plea for Intellectual Values and Social Aspects of Science

Toward the end of Chapter 2, in a one and a half page long section Sober addresses "The politics and legal status of the intelligent design hypothesis." This discussion betrays that Sober works with a philosophical framework that has been abandoned by many philosophers of science, as among other things it is blind to the philosophical relevance of intellectual values and the social context in which knowledge is generated. Sober acknowledges that intelligent design is tied to policy questions concerning the separation of church and state, with the Discovery Institute (the

intelligent design movement's political arm) having as its goals to "replace materialistic explanations with the theistic understanding that nature and human beings are created by God" and to "see design theory permeate our religious, cultural, moral and political life".⁹ Yet Sober does not view these as philosophical issues:

These policy questions have an important *psychological* component: What do policy-makers intend, and how is a policy likely to affect the minds of students and citizens generally? The subject in the present chapter, in contrast has been *logical*, not psychological. ... This division of psychological from logical questions reflects a longstanding distinction that philosophers draw; Reichenbach (1938) called it the distinction between the *context of discovery* and the *context of justification*. (p.185)

On this model stemming from logical positivism, the justification of scientific hypotheses is a rational matter, to be analyzed by philosophy. In contrast, the very generation of scientific hypotheses is not a rational process, and not to be studied by philosophy but by the empirical sciences. As Sober puts it, "pathways to new ideas – the context of discovery – are matters for psychology and sociology to understand" (p.185).

Rather than focusing on the attempt to uncover a *logical structure* of scientific knowledge, in the last three decades philosophers have increasingly come to study *actual scientific practice* in epistemological terms.¹⁰ The movement of the so-called new experimentalism argued that

⁹ From an internal document of the Discovery Institute, which was leaked to the public (<http://www.antievolution.org/features/wedge.pdf>).

¹⁰ One illustration of this is debates about reductionism in the philosophy of biology. Theory reduction is the idea that the knowledge of several scientific fields can be logically deduced from a more fundamental theory (e.g., molecular biology). Proponents of this model acknowledged that theory reduction has not been achieved yet and that it is not an aim of biologists, while arguing that theory reduction is in principle possible (Schaffner 1976). Yet the critics wondered why this should be relevant to understand biology in practice, including reductionistic explanations and methods (Brigandt and Love 2008).

experiments are not only conducted in order to test previously formulated theoretical hypotheses, but that discovery and experimental practice has its own motivations independently of theories (Hacking 1983). This has led to an abandonment of the dichotomy between the context of discovery and the context of justification as construed by the logical positivists and Sober. Some philosophers of biology have explicitly argued that discovery is a rational process that can and is to be studied by philosophy, offering some detailed accounts of scientist's reasoning involved in concrete cases of discovery in experimental biology (Darden 1991; Weber 2005). Sober is right that scientific discovery is (also) a topic for psychology and sociology, yet modern naturalistic philosophers of science do not view such empirical approaches as non-overlapping with philosophical studies (Downes 1993; Giere 1988; Hull 1988). As the reasoning involved in the generation of scientific knowledge claims is epistemologically germane to the reliability of these beliefs, Sober is wrong in implying that discovery is not a concern for philosophy.

Apart from the illicit construal of how discovery and justification differ, Sober relies on a "division of psychological from logical questions" (p.185), where the "politics and legal status of the intelligent design hypothesis" belong to the former and thus are not of philosophical relevance. Yet in addition to what "policy-makers intend" (p.185)—a (psychological) issue acknowledged by Sober—there is also the question of what the *academic proponents of intelligent design* believe and intend. Intelligent design in the sense of the doctrine that some unknown designer somehow influenced the history of living organisms is hardly endorsed by anyone. Instead, intelligent design is a strategy adopted by a part of the creationists' movement to bring creationism into America's public schools as far as possible (Forrest and Gross 2004). The academic intelligent design proponents (most of whom are not biologists) are focused on their social-political goals, and use any means—no matter how depraved—to achieve them. Speaking to one audience, they pretend that their belief system is not tied to any religious assumptions; to

other audiences they make plain that they are part of a culture war. The academic proponents of intelligent design know too well that they are unable to conduct any research supporting intelligent design and to publish in reputable peer-referred journals—in fact they do not even have enough academically looking material to keep alive the online journals set up by themselves.¹¹ As a result, the efforts of the intelligent design movement are primarily spent on publicity and political and legal campaigning. The academic intelligent design proponents occasionally publish in engineering and biological journals. While these regular papers do not offer support for the intelligent design hypothesis, their authors still advertize them to their non-academic followers as scientific arguments for intelligent design. Intelligent design proponents publicly misrepresent legitimate criticism of their views as having been removed from their academic positions.¹²

These and other facts show that apart from conducting empirical research and possessing evidence, there is a blatant difference between evolutionary biology research groups and the intelligent design community in terms of intellectual honesty. This is not just a consideration relevant to evolutionary biology. Intellectual values are vital to academic research and in fact to any intellectual endeavour. Another such value is the uptake of criticism and the free exchange of ideas. How the academic intelligent design proponents score on this point can be illustrated by how William Dembski—the theologian with PhDs in philosophy and mathematics whose probabilistic arguments against evolution Sober criticizes—runs his own intelligent design

¹¹ See <http://www.arn.org/odesign/odesign.htm> (1996–1999) and <http://www.iscid.org/pcid.php> (2002–2005, abandoned just after *Kitzmiller v. Dover Area School District* trial, which ruled the teaching of intelligent design to be unconstitutional).

¹² Compare the statements made in the ‘documentary’ *Expelled: No Intelligence Allowed* with <http://www.expelledexposed.com>.

weblog. Dembski deletes comments critical of his posts (without acknowledging this), and eventually bans any critic from commenting any further.¹³ Should a post turn out to contain serious errors or prove to backfire given his aims, Dembski edits the post (without acknowledging that changes have been made) or even delete the whole post including the corresponding discussion thread. Furthermore, upon someone criticizing one of Dembski's publications online, Dembski threatened legal action against the critique, alleging that quoting from his publication violated copyright.¹⁴ An egregious example of what intelligent design proponents will—should the opportunity arise—do with those with whom they disagree is the story of the American ecologist Eric Pianka, who had warned in a public lecture that given current human population size, epidemics are to be expected that may wipe out large parts of the human population. Dembski subsequently reported Pianka to the US Department of Homeland Security, falsely claiming that he was a bioterrorist that advocated the destruction of humanity by spreading epidemics.¹⁵

The social dimension of the intelligent design movement is of course of interest to practicing biologists, who have to rectify the public image about evolution (which has been partially distorted by intelligent design advocates) and to emphasize that their views were generated based on properly conducted research. But social aspects of science should be and nowadays often are an issue for *philosophy*. Science is a social process and generates and rationally validates knowledge based on its particular, socially embedded practices. Institutions such as societies,

¹³ Dembski's blog (now run as a group blog) is hosted at <http://www.uncommondescent.com>. His bans are recorded at <http://tinyurl.com/blogczar>. Among many other things, Dembski published several anonymous (positive) reviews for one of his books.

¹⁴ http://scienceblogs.com/goodmath/2009/11/dembski_stoops_even_lower_lega.php

¹⁵ <http://tinyurl.com/pianka>, see also <http://pandasthumb.org/archives/2006/04/texas-academy-o.html>.

journals, and the peer-review process are essential to modern science. Funding is a relevant factor for knowledge production, and decisions as to what is government agencies fund and funding by private companies necessarily blurs the line between science and the larger society. The epistemological relevance of the distribution of labour in modern science has been recognized by the field of social epistemology (Goldman 1999; Kitcher 1993). Many philosophers of science are actively working on the social dimensions of scientific knowledge (Longino 2008). Feminist approaches have pointed to persistent biases in theoretical views based on the cultural influences on science as a social institution. Helen Longino (1990, 2002) has prominently argued against a dichotomy of rational vs. social considerations and developed an alternative epistemology of scientific knowledge that views particular social factors as constitutive of scientific rationality, including mechanisms of establishing intellectual authority, publically recognized standards, diversity of perspectives, and venues for criticism. Different intellectual approaches are to be epistemologically studied and evaluated in these terms, and the above remarks made clear how intelligent design fares in terms of diversity of perspectives and venues of criticism.

Some of these issues can be illustrated by recent trends. Scientific knowledge is generally accorded a high status by society. Yet the George W. Bush government in the US was notorious for ignoring scientific advice, selectively using studies, and even muzzling government scientists in such areas as the environment, public health, and consumer protection (Mooney 2005). The Canadian government under Stephen Harper has followed suit. Just as intelligent design proponents can substantially influence public opinion by creating the false impression that there is scientific disagreement about evolution, companies may successfully evade studies that show that their products are dangerous by merely spreading some doubt about these studies (Michaels 2008). This shows that scientists face the constant need to reassert their intellectual expertise and social authority (most recently in the case of global warming). Of course 'science' is not a

monolithic whole, and current science itself faces charges of systematic research misconduct and corruption by industry and other societal interests. But this just reinforces the philosophical need to study which institutional constellations and social processes are most conducive to the production and public propagation of scientific knowledge that is as objective as possible.

Not only are such issues about epistemic values and social aspects philosophically relevant, they are also germane to debates about evolutionary theory and inference in evolutionary biology. How intelligent design proponents differ in these respects from legitimate evolutionary biologists is quite telling about the epistemic credentials of their respective *intellectual* approaches. For instance, science is not just a collection of current theories, but a dynamical process. Even if the focus is on the theoretical content of science—as in the case of Sober’s writings—understanding its past dynamics and future promise makes it necessary to take factors into account that go beyond the logical relations among theories and observations. Intelligent design is epistemically problematic not only because its current hypotheses have epistemic deficits (as Sober demonstrates), but first and foremost because no improved design hypotheses will be forthcoming. Understanding this makes it necessary to take into account aspects of intelligent design that are social and at the same time epistemologically relevant.

Continuing the quality of his previous work, Elliott Sober’s *Evidence and Evolution: The Logic behind the Science* offers a masterful epistemological treatment of scientific inference and justification in the context of evolutionary biology. Given the focus of his philosophical project, there is no need to address intellectual values and the social dimension of scientific knowledge in detail. Yet Sober’s remarks labelling the political aspects of intelligent design as psychological and thus not philosophical suggest that he views them to be of minor philosophical relevance in general or at least to an epistemological treatment of evolutionary biology—a philosophical blind spot in any case.

Acknowledgements

I thank an anonymous referee and Bruce Hunter for comments on a draft of this essay. The work on this paper was supported by the Social Sciences and Humanities Research Council of Canada (Standard Research Grant 410-2008-0400).

References

- Brigandt, I. 2007. "Typology Now: Homology and Developmental Constraints Explain Evolvability." *Biology and Philosophy* 22 (2007): 709–725.
- Brigandt, I., and Love, A.C. 2008. "Reductionism in Biology." In *The Stanford Encyclopedia of Philosophy* (Fall 2008 Edition), ed. E.N. Zalta.
<http://plato.stanford.edu/archives/fall2008/entries/reduction-biology>
- Buller, D.J. 2005. *Adapting Minds: Evolutionary Psychology and the Persistent Quest for Human Nature*. Cambridge, MA: MIT Press.
- Darden, L. 1991. *Theory Change in Science: Strategies from Mendelian Genetics*. New York: Oxford University Press.
- Dembski, W.A. 1998. *The Design Inference: Eliminating Chance through Small Probabilities*. Cambridge: Cambridge University Press.
- . 2002. *No Free Lunch: Why Specified Complexity Cannot Be Purchased without Intelligence*. Lanham: Rowman & Littlefield.
- Doolittle, W.F., and Baptiste, E. 2007. "Pattern Pluralism and the Tree of Life Hypothesis." *Proceedings of the National Academy of Sciences* 104 (2007): 2043-2049.

- Downes, S.M. 1993. "Socializing Naturalized Philosophy of Science." *Philosophy of Science* 60 (1993): 452-468.
- Elsberry, W., and Shallit, J. in press. "Information Theory, Evolutionary Computation, and Dembski's 'Complex Specified Information'." *Synthese* (in press).
- Fitelson, B. 2007. "Likelihoodism, Bayesianism, and Relational Confirmation." *Synthese* 156 (2007): 473-489.
- Fitelson, B., Stephens, C., and Sober, E. 1999. "How Not to Detect Design: A Review of William A. Dembski's *The Design Inference - Eliminating Chance through Small Probabilities*." *Philosophy of Science* 66 (1999): 472-488.
- Forrest, B., and Gross, P.R. 2004. *Creationism's Trojan Horse: The Wedge of Intelligent Design*. Oxford: Oxford University Press.
- Giere, R.N. 1988. *Explaining Science: A Cognitive Approach*. Chicago: University of Chicago Press.
- Godfrey-Smith, P. 2001. "Three Kinds of Adaptationism." In *Adaptationism and Optimality*, ed. S.H. Orzack and E. Sober. Cambridge: Cambridge University Press.
- Goldman, A. 1999. *Knowledge in a Social World*. Oxford: Oxford University Press.
- Gould, S.J., and Lewontin, R.C. 1979. "The Spandrels of San Marco and the Panglossian Paradigm: A Critique of the Adaptationist Programme." *Proceedings of the Royal Society of London B205* (1979): 581-598.
- Hacking, I. 1983. *Representing and Intervening: Introductory Topics in the Philosophy of Natural Science*. Cambridge: Cambridge University Press.

- Hull, D.L. 1988. *Science as a Process: An Evolutionary Account of the Social and Conceptual Development of Science*. Chicago: University of Chicago Press.
- Kitcher, P. 1985. *Vaulting Ambition: Sociobiology and the Quest for Human Nature*. Cambridge, MA: MIT Press.
- . 1993. *The Advancement of Science: Science without Legend, Objectivity without Illusions*. Oxford: Oxford University Press.
- Longino, H.E. 1990. *Science as Social Knowledge: Values and Objectivity in Scientific Inquiry*. Princeton: Princeton University Press.
- . 2002. *The Fate of Knowledge*. Princeton: Princeton University Press.
- . 2008. "The Social Dimensions of Scientific Knowledge." In *The Stanford Encyclopedia of Philosophy* (Fall 2008 Edition), ed. E.N. Zalta.
<http://plato.stanford.edu/archives/fall2008/entries/scientific-knowledge-social>
- Michaels, D. 2008. *Doubt Is Their Product: How Industry's Assault on Science Threatens Your Health*. Oxford: Oxford University Press.
- Mooney, C. 2005. *The Republican War on Science*. New York: Basic Books.
- Owen, R. 1843. *Lectures on the Comparative Anatomy and Physiology of the Invertebrate Animals, Delivered at the Royal College of Surgeons in 1843*. London: Longman, Brown, Green, and Longmans.
- Paley, W. 1802. *Natural Theology, or Evidences of the Existence and Attributes of the Deity, Collected from the Appearances of Nature*. London: R. Faulder.
- Quine, W.V.O. 1953. "Two Dogmas of Empiricism." *Philosophical Review* 60 (1953): 20-43.

- Quine, W.v.O. 1969. "Epistemology Naturalized." In *Ontological Relativity and Other Essays*, ed. W.v.O. Quine. New York: Columbia University Press.
- Richardson, R.C. 2007. *Evolutionary Psychology as Maladapted Psychology*. Cambridge, MA: MIT Press.
- Schaffner, K.F. 1976. "Reductionism in Biology: Prospects and Problems." In *Proceedings of the 1974 Biennial Meeting of the Philosophy of Science Association*, ed. R.S. Cohen and A. Michalos. Dordrecht: Reidel.
- Sober, E. 1988. *Reconstructing the Past: Parsimony, Evolution, and Inference*. Cambridge, MA: MIT Press.
- . 1993. *Philosophy of Biology*. Boulder: Westview Press.
- . 1999. "Testability." *Proceedings and Addresses of the American Philosophical Association* 73 (1999): 47-76.
- . 2007. "What Is Wrong with Intelligent Design?" *The Quarterly Review of Biology* 82 (2007): 3-8.
- Weber, M. 2005. *Philosophy of Experimental Biology*. Cambridge: Cambridge University Press.