

(to appear in *Proceedings and Addresses of the American Philosophical Association 1999*)

Testability

Elliott Sober*

I.

That some propositions are testable, while others are not, was a fundamental idea in the philosophical program known as logical empiricism. That program is now widely thought to be defunct. Quine's (1953) "Two Dogmas of Empiricism" and Hempel's (1950) "Problems and Changes in the Empiricist Criterion of Meaning" are among its most notable epitaphs. Yet, as we know from Mark Twain's comment on an obituary that he once had the pleasure of reading about himself, the report of a death can be an exaggeration. The research program that began in Vienna and Berlin continues, even though many of the specific formulations that came out of those circles are flawed and need to be replaced.

Philosophers of science now generally agree that confirmation theory is a central subject. No one really doubts the importance of understanding what it takes for a statement to be confirmed or disconfirmed by an observation. There also is wide consensus that the design of experiments is an important issue, not just for philosophers, but for scientists as well. If a scientist wants to test a proposition, it is important to make sure that the experiment that is carried out actually bears on the proposition in question. Sometimes it is obvious whether this is the case, but at other times, subtle issues need to be sorted out to see whether this is so. The idea that some experiments really do test a proposition, while others do not, is not controversial, nor does it deserve to be.

Matters change when the question of testability is considered. Many philosophers of science think that there can be no "criterion of testability."¹ It isn't just that philosophers have

so far failed to figure out what testability amounts to; rather, the idea is that there is no such thing. The concept of testability, like the analytic/synthetic distinction, is supposed to be a vestige of a bygone age, whose untenability we celebrate by speaking of “the demise of logical empiricism.” The question of what makes a proposition testable should be rejected, not answered.

I think this widely shared view is not just mistaken; it is *peculiar*. If it makes sense to say that an experiment does or does not test a given hypothesis, why is it suddenly misguided to ask whether any experiment *could* test the hypothesis? It’s as if chemists took the view that it is perfectly sensible to say that some things have dissolved in water while others never did, but that it makes no sense to say that some things are water-soluble while others are not. If a set of observations provides a test of a proposition because it bears relation *R* to that proposition, then a proposition is testable when it is possible for there to be a set of observations that bears relation *R* to the proposition. Testing is to testability as dissolving is to solubility. If we can understand what testing is, we also should be able to understand what testability is.

My desire to resuscitate the notion of testability does not mean that I sympathize with the *testability criterion of meaning*. The epistemological notion of testability has nothing much to do with the linguistic notion of meaningfulness. The “linguistic turn” in philosophy (Bergmann 1964) that logical empiricism helped initiate was in this instance a turn down a blind alley. We need to recover the epistemological insights that the empiricists were trying to develop by disentangling them from the extraneous linguistic lingo in which they were expressed. If a string of words has no meaning, then of course it will not express a testable proposition; for that matter, it won’t express an untestable proposition, either. However, an untestable string of words can be perfectly meaningful. Typically, we judge whether a sentence is testable by grasping what it means and then seeing whether the proposition expressed has the relevant epistemic characteristics.²

One problem that beset the testability theory of meaning was to decide which concept of possibility should be used in the claim that all meaningful sentences must be testable. Does testability mean that it is logically possible to test a sentence, or that it is nomologically possible to

do so, or that a test is feasible given current technology? This question matters to the testability theory of meaning, but it does not matter at all if we merely want to recognize testability as an important epistemological concept. It is no embarrassment that the phrase “possible to test” has multiple interpretations; there is no need to say which is the right one. Perhaps some sentences can be shown to be untestable by purely logical considerations. Others may be untestable owing just to facts about laws of nature. Still others may fail to be testable for contingent reasons – for example, because the testers we have in mind have a particular spatio-temporal location, or size, or sensory system, or because they lack some pertinent piece of information. Within this category of contingently untestable statements, we can recognize that a statement may shift from testable to untestable, or in the opposite direction, as our circumstances change. Perhaps some statements about the Kennedy assassination were testable shortly after the event took place, but became untestable subsequently. The passage of time can be an information-destroying process (Sober 1988). Conversely, statements that once were untestable can become testable, as knowledge changes and technology improves.³

Testing a hypothesis requires that it make a prediction that can be checked by observation. What, then, is an observation? Just as the idea of there being a criterion of testability has gone out of fashion, so too have many philosophers become skeptical about the concept of observation. The distinction between observational and theoretical statements is supposed to be inherently flawed, as is the distinction between observable and unobservable entities. The defects that have been advertised are various – sometimes the distinctions are said not to exist, at other times they are said to be vague, and at still other times, they are said to lack epistemological significance. Once again, I think that the negative verdicts have been exaggerated.

Here is a simple but important fact about human beings – we make observations in order to learn about things that we do not observe. This is how we learn about dinosaurs – we look at their fossil remains. This also is how we learn about quarks – we look at the measurement devices in our laboratories. The fact that dinosaurs, in a sense, are observable entities, while quarks, in a sense, are not, is irrelevant. The point is that we have *actually* observed neither. The

fact that we *could* see a live dinosaur, if we were at the right place and time, but *could not* see a quark, no matter where we went, does not show that the evidence we *actually* have supports our beliefs about dinosaurs more than it supports our beliefs about quarks (Churchland 1985). It is the distinction between observed and unobserved that pertains to questions about strength of evidence, not the distinction between observable and unobservable (Sober 1993a).⁴

The epistemological significance of this point needs to be stated carefully. It isn't that our beliefs are immune from error when they describe what we observe, but that they are vulnerable to error when they describe what we do not observe. Our opinions in both categories are fallible. Nor is it true that *all* our beliefs about the things we observe are better justified than *all* our beliefs about the things we do not observe. Rather, the asymmetry is more modest. Suppose we want to test a dinosaur fossil in order to make an inference about the dinosaur's diet. Our inference relies on the existence of a causal chain -- from *X* to *Y* to *Z*. *X* is the dinosaur's diet, *Y* is the state of the dinosaur's bones while it was alive, and *Z* is the present state of the fossil. The relationship of *X* to *Y* to *Z* is probabilistic. The dinosaur's diet does not *determine* what its bones are like, nor does the state of its bones *determine* the state of the fossil we have before us. Even so, later links in the causal chain provide evidence about earlier links. If we know the relevant present state of the fossil, this helps us infer the state of its bones when it was alive. And if we know the state of the dinosaur's bones when it was alive, this helps us infer its diet. Suppose that the chain is *singly connected* -- the only influence that *X* has on *Z* is by way of its influence on *Y*. We observe *Z* and use the information we obtain to infer the state of *Y* and also the state of *X*. In this circumstance, the following epistemological asymmetry obtains -- our knowledge of *Z* is better grounded than our knowledge of *Y*, and our knowledge of *Y* is better grounded than our knowledge of *X* (Sober 1993a). Each link in the causal chain introduces an additional source of possible error. When we infer causes from observed effects, causes that are more distal are more difficult to know than causes that are more proximate. If we *observe* the final effect in a causal chain, but not the earlier ones, then we have an epistemological asymmetry between what we observe and what we do not observe, but only infer. This asymmetry is, so to speak, "chain internal." It concerns the *X*, *Y*, and *Z* on a single causal chain; I am not comparing the more distal

causes on one chain with the more proximate causes on another.

The asymmetry just noted is a property of *singly* connected causal chains -- the only connection of *X* to *Z* passes through *Y*. When *X* and *Z* are *multiply* connected, even the modest asymmetry just noted can fail. Suppose the rock before you includes a fossilized bone, but also some other materials from the same stratum. If so, the tests you perform on the rock may provide several types of evidence that bear on the dinosaur's diet. One line of evidence might bear on what the dinosaur's bones were like when it was alive; another might reveal what plants were in the dinosaur's habitat. In this circumstance, it may turn out that your inference concerning the dinosaur's diet is stronger than the inference you draw concerning the state of its bones. With multiple connections, a distal cause can be more knowable than a cause that is more proximate.⁵

Let me describe another example in which the difference between single and multiply connected chains can be seen. Perhaps you remember the childhood game called "pass the secret" or "telephone." One child makes up a sentence and whispers it to a second, who whispers it to a third, and so on. At the end of the chain, the last child has to guess what the first child's sentence was. The process of transmission usually makes the initial sentence all but unrecognizable. In this singly connected chain, the last person has a better grip on what the penultimate person said than on what was said by the person before. Each transmission involves a new chance of distortion. For this reason, the effect at the end of the chain provides more evidence about causes that are proximate than about causes that are distal. But now imagine a different game in which the first child initiates twenty separate chains. The first child whispers in the ears of twenty children, who each go into their separate classrooms and pass the message down the line in each. The last person in each classroom then comes out and these twenty messengers each whisper in the ear of a teacher. The teacher is at the end of twenty chains, each tracing back to the initial child. She has twenty separate connections to that first child, but only a single connection to each of the intervening children. For this reason, the teacher will have a better chance of figuring out what the initial child said than of deciphering what some child at the middle of the chain in one of the classrooms said. Multiple connections are good correctives for transmissions that are subject to

error.⁶

What we know about the physical objects that we do not observe always depends on what we know about objects that we do observe.⁷ However, it isn't true that everything we know about the objects we observe depends on what we know about the objects we do not observe. Our knowledge of electrons depends on our knowledge of meter readings; but what we know about meter readings sometimes is independent of what we know about electrons. This is an important asymmetry. Again, it is important not to exaggerate. It isn't true that *everything* we know about what we observe is independent of *everything* we know about what we do not observe. *Some* of what we know about tables and cats is influenced by what we know about atoms and genes. But *some* is not. *That's* the point.

The epistemological significance of observation does not depend on whether we give the concept a broad or a narrow reading. Do scientific instruments allow us to observe things that otherwise would be unobservable, or do they merely allow us to infer states of the world from the instrument readings that we observe? As Ian Hacking (1981) once asked, do we *see* microscopic objects by looking through a microscope, or do microscopes merely allow us to infer the existence and properties of those things? There are interesting issues to address here in the philosophy of perception,⁸ but they don't matter to the epistemological issues concerning testability. Towards the end of the causal chain that culminates in an observation, there is an object that we see or hear or taste or smell or feel. The detection devices used in science are designed to connect with our sensory systems in this way. We may want to reserve our use of sensory terms like "see" and "hear" for the relations we bear to some of the last links in this causal chain, but, if we do, we must recognize that our ability to detect has a far greater reach. We can detect whether an object has a given property if we can find a causal chain whose terminus is a perceptual state we are able to enter that provides evidence as to whether the object has that property. Instrumentation allows us to detect the temperature at the earth's core, even if we can't see or hear or taste or smell or feel it. The concept of testability depends on the concept of observation, but it doesn't matter whether we equate observation with the broad notion of

detection or with the narrow notion of sensing.

The term “observation” has a sliding and context-dependent meaning in science because of the dialectical role that so-called “observations” play in resolving disputes among competing hypotheses. Observations are able to help answer a question concerning which of several competing hypotheses is most plausible only to the extent that we are able to decide which observation statements are true without first having to know which of the competing hypotheses is true. It is in this sense that observations must be “theory-neutral” – they should be neutral, relative to the competing theories under test; they need not be neutral in any absolute sense (Shapere 1982; Fodor 1983, p. 24; Sober 1994). Other theories can be used to justify various observation statements, but the theories under test cannot be presupposed by the observation statements that are used to test those very theories. We may be inclined to treat the mass of the earth as something that we observe when we use the measured mass to answer a question about something else. But if we are trying to figure out what the mass of the earth is, we may want to reserve the term “observation” for a different class of statements, ones whose truth values we can ascertain without already knowing what the mass of the earth is, and which allow us to discriminate among different estimates of that quantity.⁹

Although a hypothesis must make predictions about observations if it is to be testable, there is a second requirement that the concept of testability imposes. This is the idea that testing is an inherently *contrastive* activity -- testing a hypothesis means testing it *against* some set of alternatives (Sober 1994). The point can be seen by considering an example. At home I have a copy of Reichenbach’s 1938 book, *Experience and Prediction*. Does page 38 of that book contain an odd or an even number of letters? It would be easy to answer this question empirically; the experiences that any of us would have when we count the letters that appear on page 38 would favor one of those propositions over the other. This is because the proposition that *there are an odd number of letters on page 38* and the proposition that *there are an even number of letters on page 38* make different predictions. It therefore may seem obvious that each of those *propositions* is empirical because the *problem* of discriminating between them can be

solved empirically.

But now consider a new question: Are there an odd number of letters on page 38, or is an undetectable evil demon manipulating our experiences so that we falsely believe that there are an odd number of letters on the page? Here we have a problem that cannot be solved by making observations. The experiences we would have by counting the letters on the page would not discriminate between the two hypotheses. And apparently, there is no other empirical test that will do any better. However, in saying that the *problem* cannot be solved by an empirical test, we should not conclude that the *propositions* that enter into the problem are untestable. For the proposition that there are an odd number of letters on page 38 is a part of *both* the problems that I have described. It is part of *odd* versus *even*, which is empirical, but it also is part of *odd* versus *demon*, which is not. We don't want to say that the *odd* proposition is and is not testable. Better to say that the proposition is testable relative to one competing hypothesis, but is untestable relative to another. A question has a set of alternative answers; the problem is to see if observations can help us determine which answers are more plausible and which are less. The fundamental object that can be said to be empirical or nonempirical is a problem or question; propositions are testable only derivatively.

The position I am defending is very much in the spirit of Carnap's (1950) "Empiricism, Semantics, and Ontology," a work that I think has been widely misunderstood. Carnap, of course, would have agreed that there is a difference between questions that can be answered empirically and questions that can't. However, he additionally thought that this distinction applies to propositions, and not just to problems; he thought that there are internal and external *statements*, as well as internal and external *questions*. Carnap wanted to mark the distinction between internal and external statements by a syntactic device in a formal language. For example, if the statement that physical objects exist is external, then the way to prevent confusion about this is to use a formal language in which a special quantifier is stipulated to range over physical objects. Whether or not you share Carnap's penchant for formal languages, you should see that it is no criticism of Carnap's proposal that one's choice of quantifiers is a pragmatic matter. You

can choose to use quantifiers that range just over the physical objects, but you alternatively may want to use quantifiers that range over domains that are more inclusive, or less. This is just a matter of convenience, as Quine (1953) points out in his essay “On Carnap’s Views on Ontology.” Carnap would have agreed. It is a pragmatic matter whether you decide to mark the epistemological status of a proposition by using a language in which all propositions with that status are expressed by sentences that have a distinctive syntactic feature. It does not follow, however, that the epistemological property that Carnap was discussing is unreal, or that it is merely an artifact of our pragmatic linguistic decisions.

So far, I have suggested that an empirically soluble problem is one in which the competing hypotheses make different observational predictions. But in what sense do hypotheses “make predictions,” in view of the so-called Duhem-Quine Thesis (Duhem 1914; Quine 1953)? My answer is to separate a simple logical point that I accept from a more contentious epistemological point that I reject. The simple logical point is that hypotheses rarely make observational predictions on their own; they require supplementation by auxiliary assumptions if they are to be tested. Schematically, it isn’t the hypothesis H alone that predicts whether O will be true; rather, it is the conjunction $H\&A$ that has this implication. The controversial, and I think mistaken, epistemological point that Quine (1953) famously defended in “Two Dogmas of Empiricism” is that what gets confirmed and disconfirmed by observations is not H taken by itself, but the conjunction $H\&A$. This is Quine’s epistemological holism. It is wholes that observations impinge upon, not their parts.

The error that I see in epistemological holism is that it overlooks the fact that auxiliary assumptions are often independently tested (Giere 1988, Sober 1993c, Mayo 1996). When scientists want to test one hypothesis against another, they don’t simply *invent* auxiliary assumptions that permit the competing hypotheses to issue in predictions. Rather, they try to find auxiliary assumptions that they already have good reason to think are true. This means that the auxiliary assumptions *used* in a test and the hypotheses *under* test differ in their epistemological standing. The observational outcome favors one competing hypothesis over the others. But the

test typically will not test the auxiliary assumptions at all. For one thing, the auxiliary assumptions are independently supported; for another, scientists usually have good reason to think that the outcome of the experiment will not furnish a reason to doubt the auxiliary assumptions. Typically, the auxiliary assumptions are *epistemically independent* of the test outcome.

Consider a mundane example. A pregnancy test allows a woman to gain evidence as to whether she is pregnant. The outcome of the test is able to play this role because pregnancy tests are *reliable*. By reliability, I mean that the test has small probabilities of false positives and false negatives -- if she is pregnant, the test will probably come out positive, and if she is not pregnant, the test will probably come out negative.¹⁰ In the typical situation in which such tests are used, the outcome of the test does not change one's opinion about the test procedure's reliability. The woman taking the test doesn't know beforehand whether she is pregnant, but she *does* know that the test is very reliable. After the test results are in, she has evidence as to whether she is pregnant, but her degree of confidence in the reliability of the test procedure remains unchanged. The observational outcome tests one hypothesis against another, but it does not test the auxiliary assumptions that are used. It takes a very different experiment to test the reliability of the test procedure.

Notice that this example conforms to the unobjectionable logical point conceded before, but not to Quine's epistemological holism. Hypotheses on their own don't issue in predictions; rather, it is hypotheses conjoined with auxiliary assumptions that do so. However, the test outcome not only discriminates between the conjunction "she is pregnant and the test procedure is reliable" and the conjunction "she is not pregnant and the test procedure is reliable." In addition, the test outcome discriminates between the hypothesis that the woman is pregnant and the hypothesis that she is not. Furthermore, the test outcome says nothing about whether the test procedure is reliable. Not only are wholes tested against wholes; in addition, some of the parts are tested, but others are not.¹¹

Although this example helps illustrate where epistemological holism goes wrong, it is

unusual in an important respect. The two hypotheses that a pregnancy test allows you to test are *exhaustive*; given that Ms. X exists, there are just two possibilities -- either she is pregnant or she isn't. However, hypotheses in science rarely can be tested against their negations. Newtonian theory makes predictions, but its negation does not. Its negation subsumes all specific alternatives to the theory, both known and unknown. To know what the negation predicts, we'd have to know two things -- what each specific alternative predicts, and also what probability each of these alternatives has of being true, if Newtonian theory is false. The first of these is something that we'll never know; the second is a quantity that isn't even well-defined. Fortunately, these imponderables do not need not to be pondered. Typically, one tests two or more *specific* hypotheses against each other, where the hypotheses are not exhaustive.¹²

In saying that epistemological holism is wrong, I mean that it is wrong in the vast majority of cases. The conjunctions made of hypotheses under test and auxiliary assumptions almost always can be pulled apart and the credentials of the conjuncts evaluated separately. There may be circumstances, however, in which this cannot be done. For example, Reichenbach (1958) argued that hypotheses about the geometry of physical space and hypotheses about physical forces cannot be tested independently. The conjunction of Euclidean geometry and a physics that postulates what Reichenbach called universal forces is predictively equivalent with the conjunction of nonEuclidean geometry and a physics that denies that there are universal forces. If we assume that there are no universal forces, we can perform observational tests to determine what the geometry of space is. And if we assume that space is Euclidean (or that it is not), we can test whether there are universal forces. What we cannot do is test which conjunction of geometry plus physics is correct. My point here is not to endorse Reichenbach's argument, but to point out that his holistic conclusions are not ruled out by what I've said against epistemological holism. Reichenbach's argument needs to be evaluated by attending to specific epistemological questions about physical forces and geometry. In effect, Quine's holism is a generalization of Reichenbach's claim. It is this *generalization* that I think is mistaken.

Thus, when I say that auxiliary assumptions typically can be tested independently, I do not

mean that *all* assumptions are dischargeable. Perhaps there is a *residuum* that resists this treatment. Empirical tests are performed within the framework of a background logic. I do not claim that logic is empirically testable.¹³ And Hume claimed that all inductive inferences rest on the untestable assumption that nature is uniform (on which see Sober 1988, chapter 2). My objection to epistemological holism does not contradict this Humean thesis. I do not say that *every* auxiliary assumption that is used to test a set of competing hypotheses is independently attested by empirical evidence. I say that this is true for *many*.

I hope these comments on epistemological holism indicate how I think the following criticism of the notion of testability should be answered. It might be suggested that the concept of testability is subject to trivialization -- one is forced to conclude either that *no* statement is testable or that *all* statements are testable. The first conclusion may seem plausible if we hold that a testable statement must be able to make predictions without the mediation of auxiliary assumptions. The second conclusion may seem plausible if we hold that a statement is rendered testable merely by inventing auxiliary assumptions that allow the statement to issue in observational predictions. The way to steer between this Scylla and Charybdis is to say that testing one statement against another requires that one use auxiliary assumptions whose plausibility is attested by independent evidence.

The demand that one have independent reason to think that one's auxiliary assumptions are true leaves it open that one's auxiliary assumptions may include *idealizations*. For example, Newton frequently assumed in his calculations that the planets are point masses. This is perfectly legitimate in the context of testing a set of hypotheses, provided that the idealization does not distort the interpretation of data. What is essential is not that one be able to say that a set of assumptions is *true*, but that it is *harmless* -- that correcting the idealization would not affect the conclusions one draws. However, it isn't enough just to *assert* that the idealization is harmless; one must have *evidence* that this is so

Let's now consider another aspect of the concept of testability. If the conjunction of

hypothesis under test and auxiliary assumptions makes predictions about observational outcomes, what does “prediction” mean? The relationship is almost never deductive. All observation is subject to error, which means that the hypothesis being tested, coupled with plausible auxiliary assumptions, says which observational outcomes are more probable and which are less. The conjunction does not say which observational outcomes *must* occur and which *cannot*. The point is clearest when the competing hypotheses themselves use the concept of probability. The hypothesis that a coin is fair can be tested against the hypothesis that the coin has a probability of landing heads of, say, 0.9. These hypotheses, when supplemented with standard assumptions about the tossing process, make very different predictions about what will happen in a run of tosses. However, neither hypothesis says which observational outcome *must* occur. Each hypothesis is consistent with all possible mixes of heads and tails. The same point holds when the hypothesis under test does not use the concept of probability. A pregnancy test allows a woman to test the nonprobabilistic hypothesis that she is pregnant, but such test procedures have nonzero error rates. It is *false* that the test must come out positive if she is pregnant, and negative if she is not.¹⁴

Not only is prediction a probability concept; in addition, the relevant probabilistic question is comparative, not absolute. The probability of getting exactly 496 heads in 1000 tosses is very small, if the coin is fair. However, the probability of that outcome, if the coin’s probability of landing heads is 0.9, is much smaller. That’s why observing 496 heads in 1000 tosses favors the first hypothesis over the second. Don’t ask whether a hypothesis can or cannot explain an outcome. And don’t ask whether the hypothesis says that the outcome was probable or improbable. The relevant question is whether the outcome is *more* probable according to one hypothesis than it is according to another.

In the case of the two hypotheses about the coin, each says how probable it is that there should have been 496 heads in 1000 tosses. However, competing hypotheses need not assign precise probabilities to the observations for the hypotheses to be tested against each other. Consider, for example, a case of suspected plagiarism (Salmon 1984, Sober 1988). A teacher

notices that two student essays are virtually identical. Two possible explanations come to mind. The hypothesis of separate origination says that the two students worked independently. The hypothesis of single origination says that one student copied from the other (or that the two students copied from a common source). We know that the observed matching is far more probable on the plagiarism hypothesis than it is on the hypothesis of independent origination; yet, we are hard pressed to say exactly how probable the observed sequence of words is according to either hypothesis. However, since the point is just to *compare* the hypotheses, it suffices that we know how the probabilities provided by the two hypotheses themselves *compare*. Knowing the absolute values of the probabilities would suffice for this task, but this is not necessary.¹⁵

These various points can be summarized by a slogan: *there is no probabilistic analog of modus tollens*. If a hypothesis deductively entails something *false*, then the hypothesis is *false*. But if a hypothesis says that what you observe was very *improbable*, what then? It does not follow that the hypothesis itself is improbable. Has the hypothesis strained your credulity too much if it tells you that what you observe had a probability of only one in a zillion? Should we reject hypotheses if they step over this line that we draw in the sand? There is no such absolute threshold. The evidence points *away* from one hypothesis only in the sense that it points *towards* another. We can judge which hypotheses do better and which do worse in their competition, but that is all (Royall 1997, ch. 3).¹⁶

II.

I now have finished the “theoretical” part of my discussion of testability. Next, I want to apply these ideas to the continuing conflict between the hypothesis of intelligent design and the hypothesis of evolution by natural selection. The positivists famously held that the statement “God exists” is untestable and that the same is true of its negation. It is less well known that Popper (1974) at one time considered evolutionary theory to be unfalsifiable, and said, instead, that it is “a metaphysical research program.” With the demise of logical empiricism, both these claims have attracted much less philosophical discussion, notwithstanding the occasional attempt

by a creationist to show that evolutionary theory is a tautology and not a testable proposition at all (Bethel 1976). I think it is worth taking a fresh look at these questions.

Contemporary evolutionary theory describes a number of processes that can influence the traits that living things possess. The most famous of these is natural selection. However, it needs to be recognized that evolutionary theory includes other possible causes of evolution besides natural selection; it also needs to be recognized that natural selection is not a single process, but several.

Because the toolkit of evolutionary theory includes more than just the concept of natural selection, the theory allows for the possibility that organisms may often be *imperfectly* adapted to their environments. One reason this is possible derives from the hypothesis that there is a single *tree of life*; this is the claim that all present day organisms on Earth have common ancestors. Natural selection might favor a given trait in a lineage, but the lineage will have begun with a suite of ancestral characteristics whose influence it cannot entirely escape. Selection is often a force that tends to move lineages away from their ancestral condition, but that ancestral condition itself constitutes a force that resists the impulse to change. This influence of ancestors on descendants is sometimes called “phylogenetic inertia.” The frequent result of this conflict between selection and inertia is that the traits found in descendants are “compromises.” Consider a quantitative characteristic – say the length of a limb – that is found in a lineage. If the ancestors at the start of the lineage have a value of 5 and the optimal value for their descendants (given their environment and background biology) is 15, then it may turn out that the observed value in the descendants is between 5 and 15. If the observed value in the lineage is close to 15, that is evidence that selection was strong and that the influence of ancestors was weak; if the observed value is close to 5, precisely the opposite assessment makes sense.¹⁷

In addition to natural selection and phylogenetic inertia, contemporary evolutionary theory recognizes mutation, recombination, migration, drift, and correlations of characters induced by the underlying genetic system as possible influences on trait evolution. The theory acknowledges a

plurality of possible causes; biologists must determine empirically which combinations of causes influenced the evolution of particular traits in particular lineages. Biologists who debate adaptationism agree (or should agree) on this background theory, which describes what is *possible*. Their disagreement concerns which causes were *actual* (Sober 1984, 1993c).

With respect to the concept of natural selection, the first thing to notice is that selection doesn't simply promote the evolution of traits that are "good" in some vague sense. The theory gives this idea a precise meaning. Natural selection favors traits that promote reproductive success. Traits that promote survival will be selected only to the extent that survival is relevant to reproductive success. But now we must ask -- *whose* reproductive success will natural selection serve to enhance? Darwin almost always thought of individual organisms as the beneficiaries, though he did think that natural selection sometimes causes traits to evolve because they help groups, not individuals. In contemporary theory, these two "units of selection," as they now are called, are two items in a hierarchy. Below the individual, there are the genes that exist inside an individual. And above a group of conspecific organisms, there are multi-species communities of organisms. Evolutionary theory acknowledges that selection can occur at all these levels; empirical inquiry must determine, on a case-by-case basis, which types of selection process influenced the evolution of different traits. Group selection and individual selection predict different evolutionary outcomes, and the same is true of intragenomic conflict and community selection. Biologists who debate the units of selection problem agree (or should agree) on this background theory, which describes what is *possible*. Again, their disagreement concerns which causes are *actual* (Sober and Wilson 1998).

When biologists test evolutionary hypotheses in their scientific work, they are not testing the overarching framework that describes the possible causes of evolution. Rather, they test specific hypotheses that seek to describe why specific traits are present in specific populations. Since testing is a contrastive activity, they test specific models against other specific models. The goal is not to determine whether an evolutionary explanation can be invented that is consistent with the theory's delineation of the possible causes of evolution, but to determine whether there is

evidence that discriminates between different evolutionary hypotheses concerning what actually occurred.

As an example, consider the continuing debate in evolutionary theory about why sexual reproduction is found in many, but by no means all, organisms. One hypothesis is that sexual reproduction allows parents to produce offspring that differ among themselves, and that this strategy is advantageous when the environment is uncertain. Parents vary their offspring for the same reason that savvy investors diversify their portfolios. This hypothesis can be tested by seeing whether sexual species tend to live in environments that are more uncertain than the environments that asexual species inhabit.¹⁸ This hypothesis would be tested against the “null” hypothesis that there is no correlation between mode of reproduction and environmental uncertainty. Notice that it isn’t enough simply to *invent* a hypothesis about why sex might have evolved. One needs to extract from the hypothesis a testable prediction and then one needs to see whether that prediction is true. Those who think that evolutionary biologists simply sit around making up just-so stories about the survival of the fittest may be surprised to learn that there actually are *unsolved problems* in evolutionary theory. Various explanations have been suggested for why sex exists, but it remains controversial which of them is right.

I hope this gives a sufficiently clear picture of how evolutionary hypotheses are tested. I’ll now turn to the hypothesis of intelligent design. Creationists disagree among themselves about a number of things. Some hold that the earth is young, while others concede that it is ancient. Some think that each species was separately created by intelligent design, but others agree with Darwin’s tree of life hypothesis. They merely assert that some traits made their first appearance in life’s history because of intelligent intervention. These episodes might include the origin of life, basic features of cellular machinery, or consciousness. I set to one side the sort of theism that concedes that all features of all organisms are the result of mindless natural processes, but then insists that God set these natural processes in motion. Deism of this sort is not in competition with evolutionary theory.

If creationism is variegated in this way, what can it be said to predict about the observable features of living things? Let us begin by noting that some versions of creationism make very definite predictions. Suppose we assume that God, if he existed, would want to make all organisms *green*, and that he would have the power and the knowledge to be able to achieve this goal. This version of creationism predicts that all organisms should be green. It and Darwinian theory therefore make different predictions about what we should observe. The observations, it turns out, are squarely on the side of Darwin. Not that anyone ever defended green creationism, but the example suffices to make it obvious that there are versions of creationism that can be tested against evolutionary theory.

Not all versions of creationism have this status. Consider, for example, the hypothesis that God designed each species to have the traits it would have had if it had evolved by Darwinian processes. Call this the hypothesis of the “trickster God.” It was approximated by Philip Henry Gosse, who claimed, in his 1857 book *Omphalos*, that it was part of God’s plan of creation to put misleading fossils in the ground and a misleading navel in Adam’s belly (Gould 1985). This version of creationism agrees with the predictions that Darwinian theory generates. However, that does not mean that the trickster God hypothesis is untestable in any absolute sense. The trickster God hypothesis makes predictions just as much as Darwinian theory does -- after all, they make *the same* predictions. The trickster God hypothesis cannot be tested against Darwinian theory, but it can be tested against the green version of creationism. The trickster God hypothesis fits the observations better than green creationism does. The problem is that evolutionary theory and the trickster God hypothesis are empirically indistinguishable. I do not propose to offer a nonempirical reason for choosing between them. But neither do I conclude that Darwinism and Creationism are on an epistemic par. Evolutionary theory and Creationism are each conjunctions. We need to probe the structure of each conjunction.

The hypothesis of intelligent design can be given a strong or a weak formulation. The strong form says that *all* features of *all* living things are the result of intelligent design; a weaker formulation substitutes “some” for “all.” Let us consider a modest form of creationism that says

merely that the eye is the result of intelligent design. In what way is this version of creationism a conjunction? Well, it says that “an intelligent designer created the eye” and it also says that “if an intelligent designer creates the eye, then the eye will have such-and-such characteristics.”¹⁹

Filling in the “such-and-such” permits the conjunction to generate a testable prediction. If you assume that God would make every feature of the eye green, then this conjunction makes predictions that turn out not to be true. On the other hand, if the creationist hypothesis is filled out in accordance with the idea of a trickster God, it will predict the same features that evolutionary theory predicts.

The problem is not whether the creationist hypothesis can be filled out so that it makes predictions that accord with our observations. This is easy to do. We can always *invent* auxiliary assumptions that save the phenomena when they are conjoined with the hypothesis that the eye is the result of intelligent design. The *conjunction* is testable, but what of its *conjuncts*? Can the auxiliary assumptions be tested independently of the hypothesis? That is, can assumptions about what organisms would be like *if* they had been created by an intelligent designer be tested independently of the hypothesis *that* organisms are the result of intelligent design? I have my doubts.

It is important not to be misled here by the assumption that we know what characteristics God would have if he existed. First, we should not be parochial; we should not assume that the tenets of the religion with which we are most familiar somehow define what God must be like if such a being exists. Second, even if we assume some particular conception of what God would be like if he existed, we must check whether that conception really tells us how likely God is to produce the set of characteristics we find in the organisms in question.

The fact that testing the design hypothesis requires that we have information about the goals and abilities the designer would have, if he existed, can be seen by considering Paley’s example of the watch found on the heath. One reason we are happy with the hypothesis that the watch was produced by intelligent design is that we have no trouble believing that designers

would have wanted to produce an object with the features found in the watch and that they would have had the ability to do so. It isn't enough that the watch would have a *very low probability* of exhibiting the features we observe if it were produced by purely random natural processes; in addition, Paley must defend the positive claim that the watch would have a *higher probability* of exhibiting those features, if it were the result of intelligent design. If we knew, for example, that intelligent designers would be loathe to create objects that go tick-tock, or that are made of metal, Paley's argument about the watch would collapse.²⁰

Paley cannot be faulted for failing to consider the hypothesis of evolution by natural selection; after all, he published fifty years before *The Origin of Species* appeared. Although Paley pays some brief attention to the idea of random variation and selective retention,²¹ the real focus of his argument involves comparing the hypothesis of intelligent design with a single alternative -- the hypothesis that the adaptive contrivances of organisms originated by random natural processes. Paley was right in his claim that a complex adaptive trait has a very low probability of existing if the processes in place are purely random. As latter-day creationists have emphasized, a tornado blowing through a junkyard is unlikely to assemble a well-functioning automobile.²² But the second part of Paley's argument contains a major gap. He assumes that the adaptive contrivances that we observe would have a much higher probability of existing, if they were the result of intelligent design. I claim that Paley offers no reason to think that this is so, and that his successors have done no better.²³

The problem I am describing is made vivid by an example of Dennett's (1989, p. 285). We are shown four animals -- a laying hen, a Pekingese dog, a barn swallow, and a cheetah. The hen and the dog have many characteristics that were produced by intelligent design -- not by someone's making them from nonliving materials, but by animal breeders' modifying these organisms through the process of artificial selection. No such conscious manipulation accounts for the traits of the barn swallow and the cheetah. However, suppose we don't know this, and that we are asked to inspect these four organisms and say which features of which organisms were due to conscious design and which were due to mindless evolution. How should we set about

solving this puzzle? The risk of erroneous inference is obvious. As Dennett points out, we might make the mistake of thinking that cheetahs run fast for the same reason that greyhounds do -- they have been artificially selected to do so. And the fact that the barn swallow lives in barns might lead us to think that it was artificially selected, like the Pekingese dog, to be a pet. As for the Pekingese, it is cute and cuddly, but so are pandas -- maybe cuteness evolves without the intervention of intelligent designers.

How are we to avoid these mistakes and discern which features are the result of conscious artificial selection and which are the result of mindless evolution? This is a question that creationists as well as evolutionists should be willing to engage. Virtually all present day defenders of the design hypothesis grant that numerous features of organisms were the result of mindless evolution. They simply draw the line at certain special characteristics and claim that these special few were the result of intelligent design. So I take it that evolutionists and creationists should agree that two of the organisms we are considering were worked over quite a lot by artificial selection, and two were not. The question is how we would infer this from scratch.

Let's focus on the cheetah. The obvious reason for thinking that the cheetah's swiftness is not due to artificial selection is that cheetahs ran fast long before human beings came on the scene. Notice how information about the putative intelligent designers figures in this inference. If we had no idea when people lived, or when the ancestors of present day cheetahs started running fast, this simple argument would be blocked. But let's suppose, counterfactually, that people were around when the cheetah lineage evolved its swiftness. I think we'd still be unconvinced that cheetahs run fast because of artificial selection. After all, why would human breeders have wanted to make cheetahs swift? And would they really have had the ability to effect this change? I bet that our ancestors had more pressing problems on their minds, and, in any case, their ability to intervene in cheetah reproductive behavior would have been rather limited.

This example illustrates a general point: The problem of distinguishing the products of

artificial selection from the products of natural selection is soluble only because we know something about the goals and abilities of plant and animal breeders. If we knew nothing about these human designers, the problem would be insoluble. What holds for artificial selection carried out by human beings holds in spades for the miraculous creating done by God. If we don't know what traits organisms would probably have *if* God designed them, then we won't be able to test the hypothesis *that* God designed them. It is for this reason that the hypothesis of intelligent design cannot be tested against evolutionary theory, at least at present. To be sure, the design hypothesis figures in some *conjunctions* (e.g., the one that postulates a trickster God) that make predictions. However, that's not enough; as noted before, it isn't sufficient that one *invent* auxiliary assumptions that allow a hypothesis to predict the observations we already have in hand. The point is to test, not just this whole conjunction, but the conjunct that says, in our example, that the eye was created by intelligent design. This is the fundamental difference between the design hypothesis and Darwinism. Hypotheses in evolutionary theory, as in other areas of science, require auxiliary assumptions if they are to be tested; however, as is usually the case in science, those auxiliary assumption are independently attested.²⁴

Although the hypothesis of intelligent design is historically associated with theism, it is important to see that it has nothing essentially to do with God. The hypothesis of intelligent design is consistent with the possibility that intelligent extraterrestrials long ago seeded the earth with organisms that they crafted, just as it is consistent with life's being made by the miraculous intervention in nature of a supernatural being. Surely the latter alternative is the one that is harder to test. But the first one isn't so easy. What would an intelligent civilization in another galaxy have wanted to accomplish if they had long ago seeded the earth with life? We have no idea. Were they creating copies of the life forms that inhabited their home planet, or were they conducting an exotic experiment? Would they have constructed only a few simple organisms and then allowed the rest to evolve from them, or would they have made millions of different organisms in their factories? Was their purpose commercial, or was the seeding just for fun? If questions about extraterrestrials are difficult to answer, advocates of the design argument should not be confident that they know what characteristics God would have wanted to give to

organisms on earth if he had created them.²⁵

Creationists may be tempted to respond to this challenge simply by inspecting the life we see around us and saying that God wanted to create *that*. After all, if life is the result of God's blueprint, can't we infer what the blueprint said by seeing what the resulting edifice looks like? The answer to this question is *yes*, but that does not solve the problem. If you *assume* that God created living things, you can inspect those living things and make inferences concerning his goals and abilities. Symmetrically, if you make assumptions about God's goals and abilities, you can insure that the hypothesis of intelligent design makes predictions. But neither of these procedures cracks the nut. You can't just *assume* that God created organisms, and you also can't *assume* that if God created organisms he would have made them with such-and-such characteristics. Each of these claims must be defended by evidence. It is not legitimate to assume the one to establish the other, and then assume the other to establish the one.

The idea that God is *omnipotent* may be thought to solve this problem. Actually, the very reverse is true -- omnipotence is part of the problem, not the problem's solution. To be sure, if organisms have various characteristic, and if God is omnipotent, then *God could* have created them with those characteristics. However, the question is whether an omnipotent God *would* have created organisms as we find them. What is the probability that the vertebrate eye would have the features we see that it has, if it were designed by an all-powerful, all-knowing, and all-benevolent God? I don't claim that such a God could not have designed the eye with the characteristics it has. I also don't claim that God probably would not have produced such an eye. I claim that this probability is presently unknown, even approximately.

If omnipotence doesn't solve the problem, maybe God's benevolence holds the key. Wouldn't a benevolent God have wanted to provide organisms with the ability to see? One question that needs to be asked here is -- *which* organisms? Vision is not a biological universal. To which organisms would a benevolent God have provided this ability? But even if we restrict our attention to sighted organisms, the hypothesis that a benevolent God would have wanted to

give them the ability to see hardly begins to make contact with the data. Different groups of organisms exhibit eyes that have strikingly different designs (Dawkins 1996, ch. 5). What is the probability, if eyes were the result of intelligent design, that human beings would have eyes that exhibit one suite of characteristics, that flatworms would have eyes with a different set of traits, and that dragonflies would deploy a still different piece of machinery? What needs to be explained are the *details* of the adaptive contrivances we find. The assumption that God is all-powerful, all-knowing, and all-good is not enough for the design hypothesis to confer a probability on what we observe.

Contemporary defenders of the design hypothesis frequently assume, if only implicitly, that the design hypothesis wins by default (see, for example, Behe 1996 and Dembski 1998). They assume that if contemporary Darwinian theory cannot explain this or that observed feature of organisms, then we should accept the hypothesis that the feature was brought into being by intelligent design. There are several gaps in this line of reasoning. Even if biologists *now* cannot explain some feature, why think that future work in the Darwinian mode also will fail? And even if Darwinism is inherently unable to explain the feature, why does this show that the only possible alternative explanation is intelligent design? How do we know that no other mindless process can do the trick?

However, the defect in this argument that I'm now pointing to is different. It is misleading to say that Darwinian theory, now or in the future, "cannot explain" what we observe. The worst-case scenario for Darwinism is that the theory, with appropriate auxiliary assumptions, entails that what we observe was very improbable. However, this, by itself, isn't enough to reject Darwinism and opt for the hypothesis of intelligent design. We need to know how probable it is that the features would exist, if they were the result of intelligent design. Remember -- testing is an inherently contrastive activity. We don't test Darwinism on its own; we test it against alternatives. The question of whether Darwinism can or cannot explain what we observe is the wrong question. Instead, we need to ask a question that is comparative and a matter of degree -- does Darwinism confer a lower probability on what we observe than does the hypothesis of

intelligent design (Fitelson and Sober 1998)? *Both* hypotheses must make predictions if the observations are to help us choose between them.

By adopting the understandable tactic that the best defense is a good offense, defenders of the hypothesis of intelligent design have attacked evolutionary theory's ability to explain this or that fact about living things. Some of these criticisms are completely misguided (e.g., the blanket statement that natural processes cannot lead from disorder to order, owing to the second law of thermodynamics), but others are in the neighborhood of phenomena that are not well understood in current science. Evolutionary biologists need to answer these challenges without giving the false impression that all biological problems have already been solved. Biology is not over as a subject. But at the same time, it is important not to forget that advocates of the design hypothesis have to do more than press questions about evolutionary theory. They must develop a positive account of their own (Kitcher 1984, Pennock 1998), one in which the probabilities of adaptive features, conditional on the hypothesis of intelligent design, are not merely stipulated.

III.

Testability is important in science; understanding testability is therefore an important goal for the philosophy of science. If a problem now cannot be addressed empirically, that is an important fact, even if one cannot draw the stronger conclusion that it never will be, or never could be, solved empirically. Hume went too far when he suggested that we consign to the flames any statement that is neither testable nor a matter of definition. Since we revise our views about what is testable as we change our understanding of the empirical world, untestable problems may not remain so. They should be consigned to the back burner, not to the flames. We should keep an eye on them, and promote them to the foreground of our attention if they become empirically tractable. However, as long as they remain on the back burner, we should be very clear that science has not solved these problems *yet*.

It is important to determine whether a problem *can* be addressed empirically because

the set of inferential tools that comprise what we call “the scientific method” has the function of allowing us to evaluate competing hypotheses in the light of observations. The content of this toolbox is not fixed for all time, nor is it the same in all scientific disciplines. Still, what gets included in the toolbox at a time are the methods that scientists think should be used to interpret the bearing of observations on hypotheses. Epistemological investigation into the character of different problems is thus continuous with the scientific enterprise.²⁶ Designing an experiment is just as much a part of science as carrying out the experiment. And determining whether an experiment *can* be designed is just as much a part of science as *actually* designing it. There is more to science than the activity of running tests. Yet, I think that the scientific enterprise is directed towards the goal of bringing problems to an empirical resolution, or setting them aside when this cannot be done. Science, I am suggesting, is the art of the testable.

After the search for a criterion of testability was declared a wild goose chase, philosophers started to get interested in the Peircean idea of *inference to the best explanation* (Harman 1965, Lycan 1988, Lipton 1991). Abduction seemed a worthy successor to the failed epistemology of the logical empiricists; my impression is that it has found its way into the understanding that many philosophers now have of how good inference proceeds. First, there is the idea that *scientists* engage in inference to the best explanation when they postulate the existence of electrons and genes; but, in addition, there is the idea that *philosophers* construct inferences to the best explanation when they defend the existence of universals, or of possible worlds, or of moral facts. It is here that I want to get off the bus. “Inference to the best explanation” is an unobjectionable phrase, as long as it does not give the false impression that scientific inference allows one to discriminate between empirically equivalent theories. Scientists tend to lose interest in a question when they think its competing answers are empirically indistinguishable; they do not solve the problem by a magical invocation of the principle of inference to the best explanation (Sober 1996). I am not suggesting that philosophers should similarly lose interest. However, if a philosophical problem can’t now be solved empirically, philosophers arguing *pro* and *con* should not pretend that their arguments involve purely scientific modes of reasoning.

Three examples will illustrate the kind of mistake I have in mind. I've already mentioned evil demons and trickster gods; these, of course, are just picturesque ways of expressing the idea of *fictionalism*. A *fictionalist rewrite* of a theory *T* is a conjunction -- it says that *T* is false and that what we observe is just what we'd expect to observe if *T* were true. Let's compare a fictionalist view of a well-confirmed scientific theory -- the electron theory, for example -- and a realist interpretation of that theory, which says that the theory, interpreted literally, is true. Fictionalism and realism with respect to the electron theory are empirically equivalent. Nonetheless, it is possible to find room for the following thought: "If the electron theory were true, that would explain the observations we have made, but to say that the theory is false, but empirically adequate, provides no such explanation. Realism explains why the theory is predictively successful, but fictionalism does not." Maybe so. But it isn't a *scientific* inference to conclude from this, however tentatively, that realism is true and fictionalism is false. The problem is not empirically soluble, even if realism is more "explanatory" than fictionalism.

My second example concerns the use made of the principle of parsimony by defenders of the mind/body identity theory. Consider the materialist claim that being in pain is one and the same property as having one's c-fibers fire. Cartesian dualists deny that these two properties are identical, though they concede that the properties may be nomologically equivalent. What reason is there to prefer the identity thesis over the claim that mental and physical properties are distinct, though nomologically coextensive? Smart (1959) and Brandt and Kim (1967) answered by invoking the principle of parsimony -- the identity theory postulates fewer properties, or ultimate kinds of properties, than dualism. In a similar vein, Causey (1977) argued that the identity claim explains why mental and physical properties are nomologically coextensive; dualism seems obliged to accept this correlation as a brute fact. It may seem here that a scientific argument is being advanced in support of a metaphysical thesis. After all, isn't parsimony a consideration in scientific inference? And don't scientists construct inferences to the best explanation? The answer to both questions is *yes*, but only when the competing theories make different predictions. The materialist's inference to the best explanation may sound like science, but it isn't (Sober 1996).

My last example comes from the problem of intelligent design. Let us shift our attention from the adaptedness of organisms to large-scale features of the universe as a whole. Why are there laws of nature? Science explains some laws in terms of others. But what of the laws that scientists think are fundamental? Why are *they* true? And why are there any laws at all? It might be thought that if the universe were created by a benevolent, powerful, and knowledgeable God, that *that* would explain why there are laws. Otherwise, we apparently are forced to accept this feature of the world as a brute fact (Swinburne 1968). Maybe so. But this is not a *scientific* argument for the existence of God. Science is in the business of testing alternative hypotheses against each other, where these alternatives make different predictions. In the present example, the theistic hypothesis allegedly predicts that the universe will contain laws. But what is the competing hypothesis and what does it predict? It might be thought that if there were no God, then there probably wouldn't be laws of nature. I don't see how a probability can be assigned in this case. As far as I can see, the theistic hypothesis purports to explain a fact that no other hypothesis really engages as a problem. This is an instance of inference to the "best" explanation only in the Pickwickian sense that just one explanation has been suggested. If so, I think we have strayed from the terrain of scientific inference; we are in *terra incognita*, and the phrase "inference to the best explanation" should not reassure us.

If a philosophical problem can't *now* be solved empirically, it does not follow that the problem is *inherently* nonempirical; some philosophical problems may be like this, but others may become empirically tractable, even if they are not now. Rather than seeking to characterize a timeless difference between philosophical and scientific problems, it might be better to examine how the philosophical problems at a time differ from problems that are scientific at that time. This does not mean that there is no genuine difference between science and philosophy. Nor does it mean that philosophical problems are, at bottom, either questions of science, or nonsense. Much of what is now part of science was once part of philosophy. Different scientific disciplines branched off from philosophy at different times; the separate sciences are descendants and philosophy is the common ancestor. There is no reason to think that this branching process is

over. Future generations of philosophers will look back at us and smile indulgently, charmed by the fact that we thought that this or that problem could be solved by the methods of philosophy we now have at our disposal. Some of the innovations that will permit this change to take place will occur within philosophy, but others will be empirical results in science. The boundary between what is empirically testable and what is not will shift, but I expect that the concept of testability will remain epistemologically important.

References

- Behe, M. (1996): *Darwin's Black Box*. New York: Free Press
- Bergmann, G. (1964): *Logic and Reality*. Madison, WI: University of Wisconsin Press.
- Bethel, T. (1976): "Darwin's Mistake." *Harper's Magazine*. February, pp. 70-75.
- Brandt, R. and Kim, J. (1967): "The Logic of the Identity Theory." *Journal of Philosophy* 64: 515-37.
- Carnap, R. (1932): "Über Protokollsätze." *Erkenntnis* 3: 215-228. Translated by R. Creath and R. Nollan as "On Protocol Sentences," *Nous* (1987) 21: 457-470.
- Carnap, R. (1937): *The Logical Syntax of Language*. London: Routledge and Kegan Paul.
- Carnap, R. (1950): "Empiricism, Semantics, and Ontology." *Revue Internationale de Philosophie* 4: 20-40. Reprinted in *Meaning and Necessity*. Chicago: University of Chicago Press, 1956.
- Causey, R. (1977): *The Unity of Science*. Dordrecht: North Holland.
- Churchland, P. (1985): "The Ontological Status of Observables." In P. Churchland and M. Hooker (eds.), *Images of Science*. Chicago, IL: University of Chicago Press, pp. 35-47.
- Darwin, C. (1859): *On the Origin of Species -- Facsimile of the First Edition*. Cambridge, Ma.: Harvard University Press, 1964.
- Davis, J. (1985): *The Logic of Causal Order*. Newbury Park, CA.: Sage Publications.
- Dawkins, R. (1996): *Climbing Mount Improbable*. New York: W.W. Norton.
- Dembski, W. (1998): *The Design Inference*. New York: Cambridge University Press.

- Dennett, D. (1995): *The Intentional Stance*. Cambridge, MA: MIT Press.
- Dretske, F. (1981): *Knowledge and the Flow of Information*. Cambridge, MA: MIT Press.
- Duhem, P. (1914): *The Aim and Structure of Physical Theory*. Princeton: Princeton University Press. 1954.
- Fisher, R. (1956): *Statistical Methods and Statistical Inference*. Edinburgh: Oliver and Boyd.
- Fitelson, B. and Sober, E. (1998): "Plantinga's Probability Arguments Against Evolutionary Naturalism." *Pacific Philosophical Quarterly* 79: 115-129.
- Fodor, J. (1983): "Observation Reconsidered." *Philosophy of Science* 51: 23-43.
- Forster, M. and Sober, E. (1994): "How to Tell When Simpler, More Unified, or Less *Ad Hoc* Theories Will Provide More Accurate Predictions." *British Journal for the Philosophy of Science* 45: 1-36.
- Giere, R. (1988): *Explaining Science*. Chicago, IL: University of Chicago Press.
- Glymour, C. (1980): *Theory and Evidence*. Princeton, NJ: Princeton University Press.
- Gould, S. (1980): "The Panda's Thumb." In *The Panda's Thumb*. New York: Norton.
- Gould, S. (1985): "Adam's Navel." *The Flamingo's Smile*. New York: Norton.
- Hacking, I. (1981): "Do We See Through the Microscope?" *Pacific Philosophical Quarterly* 62: 305-322. Reprinted in P. Churchland and M. Hooker (eds.), *Images of Science*. Chicago, IL: University of Chicago Press, 1985, pp. 1325-152.
- Harman, G. (1965): "The Inference to the Best Explanation." *Philosophical Review* 74: 88-95.
- Hempel, C. (1950): "Problems and Changes in the Empiricist Criterion of Meaning." *Revue Internationale de Philosophie* 11: 41-63. Reprinted with additions and changes as "Empiricist Criteria of Cognitive Significance -- Problems and Changes." In *Aspects of Scientific Explanation and Other Essays*. New York: Free Press, 1965.
- Howson, C. and Urbach, P. (1993): *Scientific Reasoning -- the Bayesian Approach*. LaSalle, IL: Open Court, 2nd edition.
- Kitcher, P. (1984): *Abusing Science -- the Case Against Creationism*. Cambridge, MA.: MIT Press.
- Lipton, P. (1991): *Inference to the Best Explanation*. London: Routledge.
- Lycan, W. (1988): *Judgment and Justification*. New York: Cambridge University Press.

- Mayo, D. (1996): *Error and the Growth of Experimental Knowledge*. Chicago, IL: University of Chicago Press.
- Orzack, S. and Sober, E. (2000): "Adaptation, Phylogenetic Inertia, and the Method of Controlled Comparisons." In S. Orzack and E. Sober (eds.), *Adaptation and Optimality*. New York: Cambridge University Press.
- Paley, W. (1809): *Natural Theology*. London: Rivington.
- Pennock, R. (1998): *Tower of Babel*. Cambridge, MA: MIT Press.
- Popper, K. (1974): "Darwinism as a Metaphysical Research Program." In P. Schilpp (ed.), *The Philosophy of Karl Popper*. La Salle, Illinois: Open Court Press
- Quine, W. (1951): "On Carnap's Views on Ontology." *Philosophical Studies* 2: 65-72.
Reprinted in *The Ways of Paradox*, 1966, New York: Random House.
- Quine, W. (1953): "Two Dogmas of Empiricism." In *From a Logical Point of View*. Cambridge, MA: Harvard University Press. 20-46.
- Reichenbach, H. (1958): *The Philosophy of Space and Time*. New York: Dover.
- Royall, R. (1997): *Statistical Evidence -- a Likelihood Paradigm*. London: Chapman and Hall.
- Ruse, M. (1988): *But Is It Science? The Philosophical Question in the Creation/Evolution Controversy*. Buffalo, NY: Prometheus Books.
- Salmon, W. (1984): *Scientific Explanation and the Causal Structure of the World*. Princeton, NJ: Princeton University Press.
- Shapere, D. (1982): "The Concept of Observation in Science and Philosophy." *Philosophy of Science* 49: 485-525.
- Smart, J.J.C. (1959): "Sensations and Brain Processes." *Philosophical Review* 68: 141-56.
- Sober, E. (1984): *The Nature of Selection*. Cambridge, Ma.: MIT Press. 2nd edition, University of Chicago Press, 1993.
- Sober, E. (1988): *Reconstructing the Past*. Cambridge, Ma.: MIT Press.
- Sober, E. (1993a): "Epistemology for Empiricists." In H. Wettstein (ed.), *Midwest Studies in Philosophy -- Philosophy of Science*, vol.18, Notre Dame, Indiana: University of Notre Dame Press, 1993, pp. 39-61.
- Sober, E. (1993b): "Mathematics and Indispensability." *Philosophical Review* 102: 35-58.

- Sober, E. (1993c): *Philosophy of Biology*. Boulder, Co.: Westview Press.
- Sober, E. (1994): “Contrastive Empiricism.” In W. Savage (ed.). *Minnesota Studies in the Philosophy of Science (vol. 14): Scientific Theories*, Minneapolis, Minnesota: University of Minnesota Press, 1990, pp. 392-412. Reprinted in *From a Biological Point of View*. New York: Cambridge University Press, 1995.
- Sober, E. (1995): *Core Questions in Philosophy -- a Text with Readings*. Englewood Cliffs, NJ: Prentice-Hall.
- Sober, E. (1996): “Parsimony and Predictive Equivalence.” *Erkenntnis* 44: 167-197.
- Sober, E. (1999): “Instrumentalism Revisited.” *Critica* //: ///-///.
- Sober, E. (2000): “Quine’s Two Dogmas.” *Proceedings of the Aristotelian Society*.
- Sober, E., and Wilson, D. (1998): *Unto Others -- the Evolution and Psychology of Unselfish Behavior*. Cambridge, Ma: Harvard University Press.
- Swinburne, R. (1968): “The Argument from Design.” *Philosophy* 43: 199-212.
- Van Fraassen, B. (1980): *The Scientific Image*. Oxford: Oxford University Press.

Notes

*. My thanks to Robert Audi, Martin Barrett, Noël Carroll, Ellery Eells, Berent Enç, Branden Fitelson, Malcolm Forster, Clark Glymour, Daniel Hausman, Greg Mousin, Larry Shapiro, Alan Sidelle, Dennis Stampe, and Chris Stephens for comments on previous drafts.

1. See, for example, the essays by Larry Laudan and Philip Quinn criticizing Michael Ruse’s testimony at the 1981 Arkansas trial concerning whether it is constitutional for the state to require that creationism be given equal time with evolutionary theory in state schools. These, and Ruse’s responses, are collected in Ruse (1988).

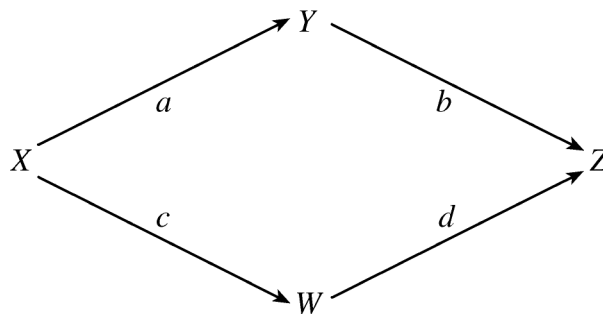
2. Hempel (1950) adopts the criterion of adequacy that if a sentence is empirically meaningful, the same must be true of its negation. Although this is a perfectly reasonable constraint on the notion of *meaning*, it isn’t obvious that it is correct as a constraint on *testability*. The constraint happens to make sense within Bayesianism, which equates confirmation with probability raising and disconfirmation with probability lowering. It is a consequence of Bayes’ theorem that $\Pr(H|O) > \Pr(H)$ iff $\Pr(\text{not-}H|O) < \Pr(\text{not-}H)$. However, nonBayesian accounts sometimes deliver a different verdict. This is the case for the ideas I’ll present in what follows.

3. Although the testability theory of meaning was a theory about language-bound entities (sentences), a purely epistemological account of testability is best formulated as a claim about language-independent entities. Hence my talk here and in what follows of statements and propositions.

4. Here I part ways with the constructive empiricism of Van Fraassen (1980).

5. Path analysis furnishes a convenient vehicle for representing the degree of correlation that obtains between nodes in a causal graph; see Davis (1985) for an elementary exposition. If there is just one path from X to Z , and it goes through Y , then the coefficient for the path is just the product of the coefficients that describe the path from X to Y and the path from Y to Z . Since these coefficients are between -1 and $+1$, it follows that the absolute value of the coefficient for Y and Z cannot be smaller than the absolute value of the coefficient for X and Z .

If there is more than one path from one node to another, then the coefficient that describes the relationship of the two nodes is the sum of the coefficients for each path. For example, suppose that X has two connections with Z -- one through Y , the other through W -- as follows:



The correlation of X and Z has the value $ab + cd$, while that between Y and Z has the value b . Notice that it is possible that $|b| < |ab + cd|$; when this is true, observing Z provides more information about X than it does about Y .

6. In this example, the teacher hears the twenty messengers; she does not hear any of the earlier children. There are values for the error probabilities that entail that she can be more certain about what the first child said than she is about what was said by any of the final twenty.

7. Here I set aside the *a priori* knowledge we have of unobserved entities. We know, for example, that an electron either has a negative charge or does not, but this knowledge is not mediated by our access to measurement devices. My point is that the *empirical* knowledge we have exhibits an asymmetry between what we observe and what we do not.

8. For example: Do we see electrons in a cloud chamber, or only their effects? Do we see the football game on live television, or only its effects? And what makes it true that you now see the printed page before you, and not your own retinal image, or the light that reaches your eyes? The general question has the following form: In a causal chain from X to Y to Z , where Z is a

perceptual state, does entering Z allow you to see (or hear) Y , or X , or both? Do we “see through” the proximate causes and thereby see causes that are more distal? Or, do we see the more proximate causes, with the distal causes occluded? See Dretske (1981) for discussion.

9. This idea was not alien to the logical empiricists; for example, Carnap (1932) uses the phrase “the relativity of protocol sentences” to make this point.

10. Notice that reliability, thus defined, does *not* mean that a woman is probably pregnant if her test is positive and is probably not pregnant if her test is negative. High values for $Pr(\text{positive test outcome} \mid \text{pregnant})$ and for $Pr(\text{negative test outcome} \mid \text{not pregnant})$ do not guarantee high values for $Pr(\text{pregnant} \mid \text{positive test outcome})$ and $Pr(\text{not pregnant} \mid \text{negative test outcome})$.

11. Glymour (1980) emphasizes the need for a nonholistic account of testing wherein auxiliary assumptions and theories under test bear very different relations to the observations.

12. This point addresses the question raised in footnote 1 -- a sentence can be testable even when its negation is not. Even if $Pr(O \mid H_1) > Pr(O \mid H_2)$, where H_1 and H_2 are incompatible, nothing follows as to whether $Pr(O \mid H_1) > Pr(O \mid \text{not-}H_1)$; this last conditional probability may not be well-defined.

13. In Sober (1993b), I argue that the mathematics used in scientific theories is typically *not* tested when those theories are tested.

14. The thesis that testing is contrastive requires that prediction be probabilistic; otherwise, hypotheses could be falsified without any contrastive alternative having to play a role. If $H \& A$ deductively entails O , and A is known to be true, then, if we observe *not- O* , we can conclude that H is false.

15. For the sake of simplicity, I am ignoring in this essay the role that parsimony plays in testing competing theories. For discussion of how parsimony interacts with the probabilistic considerations discussed here, see Forster and Sober (1994) and Sober (1996).

16. The idea that testing must be contrastive, and that each of the competing hypotheses must make predictions, has several implications with respect to statistical theory. First, it is inconsistent with the theory of R.A. Fisher (1956), on which see Howson and Urbach (1993, pp. 178-180) and Royall (1997, ch. 3); second, it raises questions about problems of statistical inference in which one of the competing hypotheses apparently makes no predictions at all; see Sober (1999) for discussion. Finally, I should mention that sometimes the fact that all the competing hypotheses at hand have low likelihoods can be a reason to search for a new hypothesis whose likelihood is higher. This is not to be construed, however, as a case in which one “rejects” the hypotheses currently available because their likelihoods are low.

17. In *The Origin*, Darwin (1859, p. 138) develops this idea when he discusses the different species of blind insects that live in dark caves around the world. These insects live in virtually identical environments, but they themselves are far from identical. Curiously, the blind insects

that live in a cave tend to resemble the sighted insects that live outside the cave nearby. The best explanation of this fact is that the insects in the cave and the insects nearby descended from a common ancestor. Here Darwin is appealing to “phylogenetic inertia” – the influence of ancestor on descendant (Orzack and Sober 2000) -- to explain imperfect adaptation. Gould (1980) makes the same argument by way of his example of the panda’s “thumb.”

18. More precisely, the question would be whether the correlation exists after one controls for other, possibly confounding, causes of trait association, such as phylogenetic inertia.

19. It is important to formulate competing hypotheses so that they do not *include* the observations. Otherwise, the observations, trivially, will be unable to distinguish among the competing hypotheses. Thus, if the creationist hypothesis is that an intelligent designer produced the eye, then the observations cannot consist merely in the fact that the eye exists. As Paley recognized, the design hypothesis has to explain the detailed features that we observe the eye to have; the same point applies to the evolutionary hypothesis.

20. To infer watchmaker from watch, you needn’t know exactly what the watchmaker had in mind; indeed, you don’t even have to know that the watch is a device for measuring time. Archaeologists sometimes unearth tools of unknown function, but still reasonably draw the inference that these things are, in fact, *tools*.

21. Paley (1809, chapter 5) presents four arguments against the hypothesis that the adaptive features of organisms arose by a process of random variation plus selective retention: (i) we do not observe the process occurring now; (ii) the hypothesis predicts that there should be many sorts of animal, now or in the past (unicorns, centaurs, etc.) that do not exist; (iii) the hypothesis falsely predicts that organisms do not exhibit a hierarchical taxonomic relationship; (iv) the claim that artifacts originate by variation plus selective retention is absurd. Paley apparently thought of the process as one-shot, rather than cumulative. Random variation assembles different combinations of matter, and the stable combinations survive while the unstable ones do not; once a combination is stable, no further variations occur in it or its descendants. Darwin, of course, thought that living things continue to experience random variation and selection, which is why he was able to claim that his theory accounts for the fact that taxonomy is hierarchical.

22. However, creationists are wrong when they say that the process of evolution by natural selection is a purely random process. Variation is generated at random, but which variants survive under selection is anything but random. See Sober (1993c) for discussion.

23. In Sober (1993c, 1995), I mistakenly said that Paley’s argument for the existence of an intelligent designer assumes a conception of God that leads to the prediction that organisms should be perfectly adapted to their environments. I am indebted to Steve Wykstra for pointing out to me that Paley should not be interpreted in this way. Paley (1809, chapters 1, 5) notes that we are right to infer the existence of a watchmaker when we see a watch, even if we notice that the watch contains imperfections. He concludes, by parity of reasoning, that the adaptive contrivances of organisms provide strong evidence that organisms are products of intelligent

design, even if these contrivances are imperfect. Paley is careful to separate the claim that an intelligent designer *exists* from the question of what *characteristics* this designer has. Paley adduces additional reasons for thinking that there is just one designer, that he is benevolent, etc., but these conclusions about the designer's characteristics do not figure as premises in Paley's argument for his existence.

24. Greg Mougin has suggested a nice example that illustrates the important contrastive element in the notion of testability. Let H_1 = God created the eye, E = Jones is pregnant, A = Jones is sexually active, and H_2 = Jones used birth control. It is possible to test H_1 against H_2 ; given independently attested background assumptions A , E favors H_1 over H_2 . The reason is that $\Pr(E | A) = \Pr(E | A \& H_1) > \Pr(E | A \& H_2)$. It also is true, of course, that E favors not- H_2 over H_1 , since $\Pr(E | A \& H_1) < \Pr(E | A \& \text{not-} H_2)$. The claim I am advancing is not that H_1 is untestable, but that H_1 cannot be tested against evolutionary theory.

25. One recent defender of the design hypothesis takes pains to point out that the designer's plans may be rather inscrutable. Behe (1996, p. 223) says that "features that strike us as odd in a design might have been placed there by the designer for a reason -- for artistic reasons, for variety, to show off, for some as-yet-undetected practical purpose, or for some unguessable reason -- or they might not." Behe therefore should agree that he has no idea whether the adaptive features we observe are ones we should expect according to the design hypothesis. What is curious is that Behe does not see this as an impediment to the design inference.

26. I hope it is clear that my reason for saying that epistemology and science are continuous in this respect differs fundamentally from Quine's. For one thing, my argument does not rest on a rejection of the analytic/synthetic distinction, which I discuss in Sober (2000). The kind of continuity I am defending here is very much in the spirit of the logical empiricists; see, for example, Carnap (1937, p. 323).